日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

EU
JP00/8133

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: １９９９年１２月１４日

出願番号
Application Number: 平成１１年特許願第３５４４０１号

出願人
Applicant (s): 日立ソフトウエアエンジニアリング株式会社

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

２０００年１２月２２日

特許庁長官
Commissioner,
Patent Office 及川耕造

出証番号　出証特２０００－３１０５５５３

【書類名】　特許願

【整理番号】　11B016

【提出日】　平成11年12月14日

【あて先】　特許庁長官　殿

【国際特許分類】　G06F 17/00

【発明の名称】　樹状図表示方法及び樹状図表示システム

【請求項の数】　13

【発明者】

【住所又は居所】　神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内

【氏名】　野崎　康行

【発明者】

【住所又は居所】　神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内

【氏名】　渡辺　恒彦

【発明者】

【住所又は居所】　神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内

【氏名】　中重　亮

【発明者】

【住所又は居所】　神奈川県横浜市中区尾上町６丁目８１番地　日立ソフトウエアエンジニアリング株式会社内

【氏名】　田村　卓郎

【特許出願人】

【識別番号】　000233055

【氏名又は名称】　日立ソフトウエアエンジニアリング株式会社

【代理人】

【識別番号】　100091096

【弁理士】

【氏名又は名称】　平木　祐輔

【選任した代理人】

【識別番号】　100102576

【弁理士】

【氏名又は名称】　渡辺　敏章

【手数料の表示】

【予納台帳番号】　015244

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9722155

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　樹状図表示方法及び樹状図表示システム

【特許請求の範囲】

【請求項１】　複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、

前記樹状図の部分木を選択するステップと、

選択された部分木を別ウィンドウで表示するステップとを含むことを特徴とする樹状図表示方法。

【請求項２】　請求項１記載の樹状図表示方法において、

前記別ウィンドウに表示された部分木に含まれる生体高分子に対するクラスタリング手法の変更を指示するステップと、

指示されたクラスタリング手法によって前記部分木に含まれる生体高分子に対して再度クラスタリング処理を行い、その結果を樹状図の形式で表示するステップとを含むことを特徴とする樹状図表示方法。

【請求項３】　複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、

前記樹状図の部分木を選択するステップと、

選択された部分木をアイコン化して表示するステップを含むことを特徴とする樹状図表示方法。

【請求項４】　請求項３記載の樹状図表示方法において、アイコン化されて表示されている部分木を元の樹状図の形式に戻して再表示するステップを含むことを特徴とする樹状図表示方法。

【請求項５】　複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、

前記樹状図の部分木を選択するステップと、

選択された部分木に含まれる生体高分子を対象として、生体高分子に関する情

報の中に予め用意されたキーワード辞書ファイルに格納されたキーワードが含まれている生体高分子の数を計数して表示するステップとを含むことを特徴とする樹状図表示方法。

【請求項６】　複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、

前記樹状図の部分木を選択するステップと、

キーワードを指定するステップと、

生体高分子に関する情報の中に指定されたキーワードが含まれている生体高分子の前記部分木内での位置を表示するステップとを含むことを特徴とする樹状図表示方法。

【請求項７】　請求項１～６のいずれか１項記載の樹状図表示方法において、前記生体高分子はｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又は遺伝子であることを特徴とする樹状図表示方法。

【請求項８】　複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するための解析を行うクラスタリング処理部と、

樹状図を表示するための表示部と、

入力手段と、

前記生体高分子に関する情報のキーワードを保持しているキーワード辞書ファイルとを備えることを特徴とする樹状図表示システム。

【請求項９】　請求項８記載の樹状図表示システムにおいて、前記入力手段によって選択された部分木を別ウィンドウで表示する機能を有することを特徴とする樹状図表示システム。

【請求項１０】　請求項９記載の樹状図表示システムにおいて、前記別ウィンドウに表示された部分木に対してクラスタリング手法を変更して再度クラスタリング処理を行い、再クラスタリング処理によって得られた樹状図を表示する機能を有することを特徴とする樹状図表示システム。

【請求項１１】　請求項８，９又は１０記載の樹状図表示システムにおいて、前記入力手段によって選択された部分木をアイコン化して表示する機能、及びアイコン化されて表示されている部分木を元の樹状図の形式に戻して再表示する機能を有することを特徴とする樹状図表示システム。

【請求項１２】　請求項８～１１のいずれか１項記載の樹状図表示システムにおいて、前記入力手段によって選択された部分木に含まれる生体高分子に対して、当該生体高分子に関する情報の中に前記キーワード辞書ファイルに格納されたキーワードが含まれている生体高分子の数を計数して表示する機能及び／又は選択されたキーワードを有する生体高分子の樹状図上の位置を表示する機能を有することを特徴とする樹状図表示システム。

【請求項１３】　請求項８～１２のいずれか１項記載の樹状図表示システムにおいて、前記生体高分子はｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又は遺伝子であることを特徴とする樹状図表示システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、特定の生体高分子、例えば遺伝子とハイブリダイズさせることによって得られたデータ（遺伝子発現データ）を、視覚的にわかりやすく、そして生体高分子（遺伝子）の機能・役割が推測しやすい形式によって表示するための表示方法及び表示システムに関する。

【０００２】

【従来の技術】

ゲノム配列が決定された種の増加に伴い、進化に対応すると見られる遺伝子を見つけ出し、どの生物にも共通に持っていると考えられる遺伝子の集合を探したり、それから逆に種に個別な特徴を推測するなど、種間の違いから何かを見出そうとする、いわゆるゲノム比較法が盛んに行われてきた。しかし近年、DNAチップやDNAマイクロアレイなどのインフラストラクチャの発達によって、分子生物学の興味は、種間の情報から種内の情報へ、すなわち同時発生解析へと移りつつあり、これまでの種内の比較と併せて、情報の抽出から関連付けの場が大きく広

がりを持ち始めている。

【０００３】

例えば、既知の遺伝子と同一の発現パターンを示す未知の遺伝子が見つかれば、それが既知の遺伝子と同様の機能があると推測できる。これら遺伝子や蛋白質そのものの機能的な意味付けは、機能ユニットや機能グループといった形で研究されている。またそれらの間の相互作用も、既知の酵素反応データや物質代謝データとの対応づけによって、あるいはより直接的に、ある遺伝子を破壊あるいは過剰反応させ、その遺伝子の発現をなくすか、あるいは多量に発現させ、その遺伝子の直接的及び間接的影響を、全遺伝子の発現パターンを調べることによって解析している。

【０００４】

この分野に成功した事例として、スタンフォード大学のP. Brownらのグループによるイースト菌の発現解析が挙げられる（Michel B. Eisen et al.: Cluster analysis and display of genome-wide expression patterns: Proc. Natl. Acad. Sci. (1998) Dec 8;95(25):14863-8）。彼らは、DNAマイクロアレイを用いて、細胞から抽出した遺伝子を時系列にハイブリダイズさせ、遺伝子の発現の度合い（ハイブリダイズした蛍光シグナルの輝度）を数値化した。そしてこの数値に応じて、細胞の一連のサイクルで発現パターンの過程が近い遺伝子どうし（任意の時点での発現の度合いが近いものどうし）をクラスタリングしている。

【０００５】

図１は、この方式にそって遺伝子の発現パターンの類似性を表現した表示例である。右側には観測した個々の遺伝子の情報が列挙されており、左側にはこれらの遺伝子の発現パターンに応じて作成された樹状図が示されている。樹状図は、クラスタリングの過程で、最も近い2つのクラスタ毎に併合されてきた状況を表しており、各枝の長さは併合時の2つのクラスタ間距離（クラスタ間の非類似度）に対応している。このような表示方法をとることで、共通のクラスタに属する遺伝子は、共通の機能的性質をもつ可能性があると推測することができる。

【０００６】

【発明が解決しようとする課題】

実際の遺伝子発現パターンの分析では、大量のデータをクラスタリングすることになる。通常、DNAチップやDNAマイクロアレイは、数千から数万の遺伝子を同時に観測することが可能である。一般に遺伝子の発現過程は、ある遺伝子の発現が別の遺伝子の発現を誘導したり、あるいは、発現を阻害するなど、遺伝子間で複雑なネットワークを形成している。それ故、観測する遺伝子の数が多ければ、より複雑で詳細なネットワークを調べることができる。

【０００７】

ところが、遺伝子の数が膨大になると、全体の遺伝子の働きを把握することは非常に困難になる。すなわち、樹状図には数千～数万の遺伝子が並ぶことになるので、この表示から、どのような分類ができているのかを判断するのは難しい。また、クラスタリング手法の違いにより、樹状図の枝の長さは一般的に異なる。例えばクラスタ併合アルゴリズムとして、最長距離法を選択したとき、枝の長さの平均は、最短距離法を選択したときの長さの平均よりも長い。したがって、樹状図全体としてみたとき、図２のように、根から葉までの長さもまた、クラスタリング手法によって異なる。遺伝子の発現データに対するクラスタリングでは、枝の長さよりも、どのように分類されているのかを調べることが重要である。それゆえ、通常、樹状図の表示を行なうときは、図３のように、樹状図の根から葉までの長さを一定値に定め、各枝の長さは根から葉までの長さに対する相対的な長さで表し、クラスタリング手法に応じて枝の長さの縮尺を変更して表示する。

【０００８】

ここで、上記のような樹状図の表示方法を採用したとき、樹状図の中に発現パターンが類似している遺伝子を多数含んでいると、枝の長さが小さい樹状図が形成されるが、これらの枝の長さが樹状図の根から葉までの長さに比べて非常に小さいと、図４の４０１に示すように遺伝子間の枝の詳細な関係を知るのが非常に困難になる。また、従来の遺伝子発現解析に関するクラスタリングでは、部分木を選択し、これに対して別のクラスタリング手法を適用するなど、対話的な操作ができなかった。また、従来の遺伝子発現解析に関するクラスタリングでは、分類が成功しているかどうかを調べる手段として、遺伝子の機能や遺伝子名のキーワードに着目し、それらが部分木に集まっているかどうかによって判断していた

。しかし、解析する遺伝子の数が膨大なものになると、どのような機能やキーワードに着目すべきかを判断するのは、非常に困難な作業である。

本発明は、このような従来技術の問題点に鑑み、樹状図全体の枝の状態を大域的に把握でき、かつ個々の部分木の状態を詳細に知ることができるような樹状図表示方法及び樹状図表示システムを提供することを目的とする。

【０００９】

【課題を解決するための手段】

上記目的を達成するために、本発明では、樹状図の枝を選択し、選択した枝から葉の部分木に対して、別の表示ウィンドウで表示する機能、アイコン化する機能、アイコン化したものを元に戻す機能、部分木に含まれるキーワードを収集し表示する機能、を備えた樹状図表示システムを提案する。本発明によると、作成された樹状図の部分木に対して、異なるクラスタリング方法を対話的に適用する処理を実現することができる。また、クラスタリングが成功しているかどうかを判別するため、部分木にどのようなキーワードが多く含まれているかを表示し、分類の絞り込みや、クラスタリング方法の選択の支援を行うことができる。

【００１０】

以下、理解を容易にするため、本発明を遺伝子のクラスタリングに適用した場合を例にとって、本発明の樹状図表示システムによる樹状図の表示例について説明する。ただし、本発明は遺伝子のクラスタリングにのみ適用されるものでなく、他の生体高分子、例えばｃＤＮＡ、ＲＮＡ、ＤＮＡ断片等についても同様に適用可能である。

【００１１】

図５は、本発明の樹状図表示システムによる樹状図の表示例を示している。分類アルゴリズムの選択メニュー５０１、及び（非）類似度の選択メニュー５０２を備えている。遺伝子発現データを読み込み、分類アルゴリズム及び（非）類似度を選択すると、樹状図が作成される。また、本システムは、図１のように遺伝子名などの遺伝子情報を樹状図の葉の先に付加して表示する形式も選択できる。

【００１２】

作成された樹状図において、任意の枝を選択すると、選択した枝から葉までの

部分木に対する操作、すなわち、この部分木を別のウィンドウで表示する、この部分木をアイコン化する、この部分木のアイコンを元に戻す、この部分木に含まれる単語を検索する、というメニューが選択できる。図は、画面中央の枝５０５を矢印で図示されているマウスカーソル５０４等で選択した状態を示しており、このとき開くメニューウィンドウ５０３には選択可能なメニューが表示されている。マウスカーソル５０４をメニューウィンドウ５０３内に移動して、所望の処理項目をクリックすると選択された処理が実行される。

【００１３】

分類アルゴリズムは、図５の状態ではウォード法が選択されているが、選択メニュー５０１を開くことによって例えば、最短距離法、最長距離法、群平均法、重心法、メディアン法、可変法など他のアルゴリズムを選択することができる。（非）類似度は、個体間の類似の程度を表す指標である。この指標には、距離のように値の小さい方が類似性が高いことを表す場合と、相関係数のように値の大きい方が類似性が高いことを表す場合がある。前者の指標を非類似度、後者の指標を類似度という。図５の状態では非類似度としてユークリッド距離が選択されているが、選択メニュー５０２から他の（非）類似度、例えば標準化ユークリッド平方距離、マハラノビスの（汎）距離、ミンコフスキー距離等を選択することができる。このとき、分類アルゴリズムとして重心法、メディアン法、可変法を選択したとき、非類似度としてユークリッド平方距離以外に選択できないなど、分類アルゴリズムと非類似度との組み合わせが妥当なものである必要がある。

【００１４】

図６は、図５に示した表示画面において、「部分木を別のウィンドウで表示する」メニューを選択したときの表示例である。図６に示すように、選択した部分木を、根から葉までの長さに応じて縮尺を変更し表示し直す。このような表示手法をとることで、利用者は部分木の詳細な枝の状態を調べることが出来る。また、本システムでは、選択した部分木に対して、分類アルゴリズム及び／又は（非）類似度を選択して、再度クラスタリングを行なうことが出来る。このようにすることで、例えば、はじめのクラスタリング結果からクラスタ間の距離が大きいもの（図４において、４０１と４０２、４０１と４０３の関係）を見つけ出し、

これを除外して、興味のある部分木のみ詳しく調べることが出来る。分類アルゴリズム及び／又は（非）類似度の選択は、分類アルゴリズムの選択メニュー５０１、及び（非）類似度の選択メニュー５０２によって行う。

【００１５】

図７は、図５に示した表示画面おいて、「部分木をアイコン化する」メニューを選択したときの表示例である。部分木５０５を７０１のようにアイコンにすることで、樹状図の大域的な状態を容易に知ることが出来る。例えば、同様の機能をもつ遺伝子群や、発現がほとんど観測されなかった遺伝子群を一つのアイコンとしてまとめるなどの利用法が可能である。

【００１６】

図８は、図５に示した表示画面おいて、「部分木に含まれる単語を検索する」メニューを選択した時の表示例である。この機能を適用すると、選択した部分木に含まれる遺伝子の中で、遺伝子に対応する遺伝子情報に予め定めたキーワードが含まれるものを数え上げ、検索結果８０１として表示する。更に検索結果８０１から、マウスカーソル８０４等で一つのキーワード８０２を選択すると、そのキーワード（図の場合、"ribosomal"）を持つ遺伝子を、マーク８０３等によって樹状図上の位置として表示する。これにより、選択した部分木にどのような遺伝子が集まっているかを容易に知ることができる。また、この結果、分類がうまくいっていないのであれば、別の分類アルゴリズムや（非）類似度を選択して再度クラスタリングを行なうなど、より適切なクラスタリング方法の選択の支援をすることができる。

このように、本発明によると、作成された樹状図から、効果的に意味を抽出することができる。

【００１７】

すなわち、本発明による樹状図表示方法は、複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、前記樹状図の部分木を選択するステップと、選択された部分木を別ウィンドウで表示するステップとを含むことを特徴とする。

本発明は、別ウィンドウに表示された部分木に含まれる生体高分子に対するクラスタリング手法の変更を指示するステップと、指示されたクラスタリング手法によって前記部分木に含まれる生体高分子に対して再度クラスタリング処理を行い、その結果を樹状図の形式で表示するステップとを含んでもよい。

【００１８】

本発明の樹状図表示方法は、また、複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、前記樹状図の部分木を選択するステップと、選択された部分木をアイコン化して表示するステップを含むことを特徴とする。

必要により、アイコン化されて表示されている部分木を元の樹状図の形式に戻して再表示するステップを含むこともできる。

【００１９】

本発明による樹状図表示方法は、また、複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、前記樹状図の部分木を選択するステップと、選択された部分木に含まれる生体高分子を対象として、生体高分子に関する情報の中に予め用意されたキーワード辞書ファイルに格納されたキーワードが含まれている生体高分子の数を計数して表示するステップとを含むことを特徴とする。

【００２０】

本発明による樹状図表示方法は、また、複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するステップと、前記樹状図の部分木を選択するステップと、キーワードを指定するステップと、生体高分子に関する情報の中に指定されたキーワードが含まれている生体高分子の前記部分木内での位置を表示するステップとを含むことを特徴とする。

上記樹状図表示システムにおいて、生体高分子はｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又は遺伝子とすることができる。

【００２１】

本発明による樹状図表示システムは、複数種類の生体高分子に対して複数の異なる条件で実験を行って得られたデータの組に基づいて前記複数の生体高分子のクラスタリング処理を行い、その結果を樹状図の形式で表示するための解析を行うクラスタリング処理部と、樹状図を表示するための表示部と、入力手段と、生体高分子に関する情報のキーワードを保持しているキーワード辞書ファイルとを備えることを特徴とする。入力手段は、樹状図の枝の選択や、クラスタリング手法の選択などに用いられるもので、例えばキーボードやマウスとすることができる。キーワード辞書ファイルは、クラスタリングの結果に対し利用者が望む形になっているかを判断するために用いることができる。

【００２２】

この樹状図表示システムは、入力手段によって選択された部分木を別ウィンドウで表示する機能を有することができる。また、別ウィンドウに表示された部分木に対してクラスタリング手法を変更して再度クラスタリング処理を行い、再クラスタリング処理によって得られた樹状図を表示する機能を有することができる。

この樹状図表示システムは、入力手段によって選択された部分木をアイコン化して表示する機能、及びアイコン化されて表示されている部分木を元の樹状図の形式に戻して再表示する機能を有することができる。

【００２３】

この樹状図表示システムは、入力手段によって選択された部分木に含まれる生体高分子に対して、当該生体高分子に関する情報の中にキーワード辞書ファイルに格納されたキーワードが含まれている生体高分子の数を計数して表示する機能及び／又は選択されたキーワードを有する生体高分子の樹状図上の位置を表示する機能を有することができる。

本発明の樹状図表示システムにおいて、前記生体高分子はｃＤＮＡ、ＲＮＡ、ＤＮＡ断片又は遺伝子とすることができる。

【００２４】

【発明の実施の形態】

以下、図面を参照して本発明の実施の形態を説明する。以下では、遺伝子のクラスタリングを例にとって説明するが、本発明の適用範囲は遺伝子のクラスタリングのみに限定されるわけではなく、ｃＤＮＡ、ＲＮＡ、ＤＮＡ断片など生体高分子一般に対して同様に適用することができる。

【００２５】

図９は、本発明による樹状図表示システムの一例を示す構成図である。このシステムは、遺伝子の情報及び発現過程を記録した遺伝子データ９０１と、遺伝子の発現過程に応じてクラスタリングを行ない、それを樹状図の形式で表示するための解析を行なうクラスタリング処理部９０２と、樹状図を表示するための表示装置９０３と、樹状図の枝や、クラスタリング手法の選択などに用いるキーボード９０４及びマウス９０５等の入力手段と、クラスタリングの結果に対し利用者が望む形になっているかを判断するための遺伝子情報のキーワードを保持しているキーワード辞書ファイル９０６から構成される。このクラスタリング処理部９０２は、コンピュータとそのプログラムによって具体化されるものである。なお、記憶装置９０１に代えて、ネットワーク等を介して遠隔地に設置されたサーバコンピュータが管理しているデータベースから遺伝子データを取得する構成をとってもよい。

【００２６】

図１０は、遺伝子データ９０１に格納された遺伝子発現パターンデータの具体的な構造を示したものである。本アルゴリズムでは、これを２次元配列によって格納する。すなわち、遺伝子ID（id）をもつ遺伝子が実験ケース（no）における発現の度合い（ハイブリダイズした蛍光シグナルの輝度）を数値化したデータを、Exp[id][no]に格納する。m種類の遺伝子をそれぞれ異なる位置にスポットしたDNAチップから得られる１回の実験は、１つの実験ケースに対応する。

【００２７】

図１１は、遺伝子データ９０１に格納された遺伝子に関する情報を格納するための、遺伝子情報構造体の例を示している。この遺伝子構造体は、遺伝子ID（１１０１）、遺伝子のORF（１１０２）、遺伝子名（１１０３）、遺伝子の機能（１１０４）のメンバから構成される。図１１はあくまでも説明のための例であり

、ここに示した遺伝子の属性以外の情報も、遺伝子情報構造体のメンバとして定義することももちろん可能である。

【００２８】

図１２は、クラスタリング処理において利用するクラスタを表す構造体の例を示している。全てのクラスタ構造体は、樹状図の各ノードまたは葉と対応している。クラスタ構造体は、ウィンドウ単位で管理され、同じウィンドウのノードまたは葉であれば、同一のwindowID（１２０７）をもつ。また、同じウィンドウ内でノードまたは葉を識別するため、clusterNo（１２０５）で各クラスタ構造体に一意に番号を割り振っている。クラスタ構造体には3種類あり、type（１２０１）の値が、leafのもの、nodeのもの及びiconのものに分かれる。

【００２９】

各leaf型クラスタ構造体は、それぞれひとつのgeneID（１２０６）に対応している。すなわち、ひとつの遺伝子に対応している。更にgeneIDから、遺伝子情報構造体のデータが参照できる。node型クラスタ構造体は、クラスタリングにおける併合処理において逐次生成するもので、併合前の2つのクラスタをleft（１２０２）の値と、right（１２０３）の値からたどれるようにし、また、それらの間の距離（（非）類似度）をdistance（１２０４）の値として保持する。left及びrightの値には、クラスタを一意に示すclusterNo（１２０５）が入っている。icon型クラスタ構造体は、部分木をアイコンに置き換えるときに生成され、表示では葉の場合と同様に扱う。そして枝の先端には部分木を示すアイコンを付して表示する。実際の部分木のルートのクラスタは、left（１２０２）の値からたどることができる。

【００３０】

図１３は、図１２に例示したクラスタ構造体のデータ構造を示した図である。これらはクラスタ分析の過程で生成される。クラスタ構造体は、最初leaf型のものだけを用意するが、クラスタリングの過程で2つずつ併合し、その度にnode型クラスタ構造体を生成してトリー構造を組み立てる。node型クラスタ構造体には、併合した2つの子ノードのclusterNoと、それらの間の距離（（非）類似度）の情報が登録されている。またleaf型クラスタ構造体に登録されているgeneIDによ

り、対応する遺伝子情報を参照することができる。アイコン化する処理があれば、トリーの途中にicon型のクラスタを挿入し、あたかも葉であるかのように表示する（表示に当たっては、icon型のクラスタより先に位置するクラスタは表示しない）。アイコンを解除するときは、icon型クラスタの上下のクラスタのリンクを繋ぎ直す操作を行う。

【００３１】

図１４は、クラスタ分析の過程でクラスタ間の距離である非類似度を格納するための配列の例である。図に示すように、2次元配列dist[ ][ ]を用いてこれを格納する。また、２次元配列dist[ ][ ]のインデックスの数字に対応するクラスタのclusterNo（１２０５）を格納した配列を、clust_idx[ ]に格納する。すなわち、非類似度dist[i][j]の値は、clusterNoがclust_idx[i]とclust_idx[j]であるクラスタ間の値を示す。図１４から、例えばclust_idx[3]であるclusterNo:9のクラスタとclust_idx[4]であるclusterNo:25のクラスタ間の非類似度dist[3][4]の値は２１であることが分かる。

【００３２】

図１５は、各ウィンドウのルートノードを格納する配列の例を示している。すなわち、各表示ウィンドウに対するルートノードのクラスタのclusterNoは、配列RootNode[ ]に格納される。図１５に示した例では、RootNode[1]の値が569であることからwindowID:1の表示ウィンドウに表示される樹状図のルートノードはclusterNo:569のクラスタであることが分かり、RootNode[2]の値が312であることからwindowID:2の表示ウィンドウに表示される樹状図のルートノードはclusterNo:312のクラスタであることが分かる。

【００３３】

図１６は、検索の問合せ及び結果を格納するためのsearch構造体の例を示している。キーワード辞書ファイル９０６に登録されている各キーワードに対して、構造体を一つ生成する。また、キーワードで同義語のものがいくつか存在するとき、それらをひとつのものを指すこととして扱うこともできる。search構造体は、検索項目のキーワードを入力しておくkeyword（１６０１）、そのキーワードが部分木の中でいくつあったかを示すtimes（１６０２）、キーワードが遺伝子

情報の中にあったとき、その遺伝子の樹状図上の位置を格納するplace（１６０３）をメンバとしてもつ。図１６に図示する例のように、Rat、Mouse、Musのような同義語をまとめてkeywordメンバに登録しておくことで、これら3つのキーワードのどれをも同じ検索項目として扱うことが出来る。

【００３４】

図１７は、本システムの概略フローを示した図である。

まず、遺伝子データ９０１からクラスタリング処理部９０２へデータを読み込む（ステップ１７０１）。これについては、後で詳しく説明する。次に、クラスタ分析、及び結果表示に必要な各種パラメータを設定する（ステップ１７０２）。ここでは、分類アルゴリズム及び（非）類似度の設定、個々の遺伝子情報を表示するか否かなどの設定を行う。

【００３５】

次にクラスタ分析を行い（ステップ１７０３）、結果を表示する（ステップ１７０４）。クラスタ分析については、後で詳しく説明する。このクラスタ分析の処理の中で、樹状図表示に必要な情報を収集し、クラスタ構造体に入力する。分析結果表示では、このクラスタ構造体と、個々のウィンドウのルートノードのclusterNoを表すRootNode[ ]の情報をもとに、結果を表示する。クラスタ構造体のtypeがiconのときは、それを葉のように扱い、枝の先端に部分木を表すアイコンを付加する。

【００３６】

表示された樹状図の中のある部分木をアイコン化してまとめる、あるいはアイコン化を解除して元の部分木に戻す場合、以下の処理を実行する（ステップ１７０５）。すなわち、樹状図の枝をマウスで選択し（ステップ１７０６）、部分木のアイコン化、または非アイコン化処理を行う（ステップ１７０７）。アイコン化、非アイコン化処理に関しては、後で詳しく説明する。処理の後、再び分析結果表示（ステップ１７０４）を行う。

【００３７】

表示された樹状図に対して、キーワード辞書ファイル９０６に格納されたキーワードをもとに検索を行う場合、以下の処理を実行する（ステップ１７０８）。

すなわち、樹状図の枝をマウスで選択し（ステップ１７０９）、検索処理を行う（ステップ１７１０）。検索処理に関しては、後で詳しく説明する。検索処理１７１０で、表示に必要な情報がsearch構造体に格納されるので、それをもとに新たに検索結果ウィンドウを生成し結果を表示する（ステップ１７１１）。このとき、マウスなどで検索結果ウィンドウのあるキーワードを選択すると、search構造体のplaceメンバの情報をもとに、樹状図上のキーワードのある箇所にマーカーを付与する。

【００３８】

表示された樹状図に対して、他の併合アルゴリズム、（非）類似度で再びクラスタリングを適用したいときは、ステップ１７０２に戻る（ステップ１７１２）。　クラスタ併合アルゴリズムとしては、例えば、最短距離法、最長距離法、群平均法、重心法、メディアン法、ウォード法、可変法等がある。最短距離法、最長距離法、群平均法、ウォード法、可変法には、次々にクラスターを融合していくときの非類似度が単調に大きくなる特性がある。また、２つのクラスターを融合して１つのクラスターを作ると、他のクラスターとの距離が近づく場合と遠ざかる場合があり、前者を空間の収縮、後者を空間の膨張、距離が変わらない場合を空間の保存と呼ぶが、最短距離法は空間が収縮する特性を有し、最長距離法やウォード法は空間が膨張する特性を有する。また、群平均法、重心法、メディアン法は、空間が保存され、可変法の場合はパラメータの設定によっていずれにもなりうる。（非）類似度にも種々のものがあり、例えば非類似度の代表的なものとしてはユークリッド平方距離、標準化ユークリッド平方距離、マハラノビスの（汎）距離、ミンコフスキー距離等がある。従って、前述の特性等を勘案して、これらの中から適宜のものを選択すればよい。

【００３９】

表示された樹状図に対して、ある部分木を別のウィンドウで表示させたい時（ステップ１７１３）は、別ウィンドウに表示したい樹状図の枝をマウスで選択し（ステップ１７１４）、選択した樹状図の部分木に対するデータの読み込みを行い（ステップ１７１５）、再びステップ１７０２に戻る。選択した樹状図の部分木に対するデータの読み込み処理については、あとで詳しく説明する。

以上の選択が無かった場合には、処理を終了する。

【００４０】

図１８は、図１７における遺伝子データの読み込み処理１７０１の詳細フローである。

まず、遺伝子数、実験ケースの総数をそれぞれgene_num、exp_numに登録する（ステップ１８０１）。次に、遺伝子データ９０１から遺伝子情報を読み取り、遺伝子情報構造体gene_info[i](i = 1,…,gene_num)に登録する（ステップ１８０２）。遺伝子データ９０１から遺伝子発現データを読み取り、Exp[i][j](i = 1,…,gene_num, j = 1,…,exp_num)に登録する（ステップ１８０３）。樹状図の葉の総数を表すleaf_numにgene_numを代入する（ステップ１８０４）。

【００４１】

次に、初期値となるleaf型クラスタ構造体を生成する。クラスタ構造体clusterをleaf_num個生成し、i = 1,…, leaf_numに対して、typeをleafに、clusterNoをiに、 geneIDをiに、 windowIDを1として登録する（ステップ１８０５）。次に、キーワード辞書ファイル９０６に格納されたキーワードを読み出し、それぞれのキーワードに対してsearch構造体を生成し、キーワードをsearch[].keywordに登録する（ステップ１８０６）。キーワードの総数をkey_numに代入する（ステップ１８０７）。windowIDを表すwidに１を登録し（ステップ１８０８）、処理を終わる。

【００４２】

図１９、図２０は、図１７におけるクラスタ分析処理１７０３の詳細フローである。

windowIDがwidに対応するウィンドウ内の遺伝子間の発現度の非類似度を求める。clusterNoがi,jに対応する遺伝子の非類似度をdist[i][j]に登録する（ステップ１９０１）。本アルゴリズムでは、クラスタが１つ生成されるごとにclusterNoを１から順に割り振っている。そこで、次のクラスタが生成されたとき、そのクラスタの番号を表すnewclusterNoにleaf_num + 1を代入しておく（ステップ１９０２）。また、クラスタ間距離（非類似度）を格納する配列の情報として、併合対象クラスタ数を示すall_clustにleaf_numを代入し、i = 1,…,leaf_numに

対し、cluster_idx[i]にiを代入して初期化しておく。併合対象クラスタの数all_clustが１に等しいかどうか判定し、等しくない場合、１になるまで以下の一連の処理を繰り返す（ステップ１９０５）。

【００４３】

最初に、先に求めたクラスタ間距離（非類似度）から、次に併合されるべきクラスタを決定する。すなわち、i ＜ j かつi，j = 1,2,…，all_clustに対して、dist[i] [j]の最小値、最小値を与えるi、最小値を与えるjを求め、d_min、i_min、j_minにそれぞれ代入する。clusterNoが、cluster_idx[i_min]、cluster_idx[j_min]のクラスタが次に併合されるべきクラスタとなる。clusterを新規に生成し、typeにnode、leftにcluster_idx[i_min]、rightに cluster[j_min]、distanceにd_min、clusterNoにnewclusterNo、windowIDにwidを登録していく（ステップ１９０７）。ここで、２つのクラスタのどちらをleftメンバとし、残りをrightメンバとするかについては、発現量で比較するなど予め判定基準を設ける方式をとることも可能である。

【００４４】

次に、クラスタ間距離を格納している配列の情報を更新する。まず、新しく生成したクラスタと他のクラスタとの距離（（非）類似度）を求め、それをi_minのクラスタと他のクラスタ間の距離が格納されていたdist[] []の配列位置に上書きする。i = 1,2,…,i_min－1に対し、新しく生成したクラスタと、clusterNoがcluster_idx[i]に対応するクラスタとの非類似度をdist[i] [i_min]に登録し、j = i_min + 1, …, j_min_1, j_min + 1, … , all_clustに対し、新しく生成したクラスタと、cluster_idx[j]に対応するクラスタとの非類似度をdist[i_min] [j]に登録する（ステップ２００１、２００２）。

【００４５】

次に、j_minに関する情報を削除して、j_min以降のすべての配列データを一つ前に移動する処理を行なう。i = min_j,…,all_clust－1に対し、clust_idx[i]にclust_idx[i +1]を代入する（ステップ２００３）。次にi ＜ j、i，j = j_min, …, all_clustを満たすi ,jに対し、dist[i] [j]にdist[i + 1] [j]を代入し、その後i＜ j、i = 1,…, all_clust－１、j = j_min,…,all_clust－１を満たす

i ,jに対し、dist[i][j]にdist[i][j + 1]を代入する（ステップ２００４、２００５）。

【００４６】

最後に、併合対象クラスタ数を示すall_clustから１を引き、新しいクラスタ構造体に割り振るclusterNoを表すnewclusterNoに１を加える（ステップ２００６，２００７）。

以上の操作をall_clustが１になるまで繰り返す。all_clustが１になれば、RootNode[wid]に、このウィンドウのルートノードのclusterNoを表すcluster_idx[1]を代入し、処理を終える（ステップ１９０８）。

【００４７】

図２１は、図１７におけるアイコン化する、または（非）アイコン化（アイコンを解除）する処理１７０７の詳細フローである。

６において選択した枝の両端に対応するクラスタを登録する。下（leaf側）のclusterをchildClustに代入し、枝の上（root側）のclusterをparentClustに代入する（ステップ２１０１，２１０２）。次に、新しくicon型clusterを生成し、childClustとparentClustの間に挿入する処理を行なう。すなわち、clusterを生成し、typeにiconを、leftにchildClust.clusterNo、をclusterNoにnewclusterNoを、windowIDにwidをそれぞれ登録する（ステップ２１０３）。そして、ポインタの付け替え操作として、parentClust.leftまたはparentClust.rightに登録されているchildClustのclusterNoをnewclusterNoに変更する（ステップ２１０４）。全体のクラスタ数がひとつ増加したので、新しいクラスタ構造体に割り振るclusterNoを示すnewclusterNoに１を加えて処理を終了する。（ステップ２１０５）

【００４８】

また、部分木をアイコン化したものを元に戻すメニューを選択すると、まず図１７におけるステップ１７０６で選択した枝の両端に対応するクラスタを登録する。ステップ１７０６で選択した枝の下（leaf側）にあるアイコンのcluster、アイコンの親ノードのclusterをそれぞれiconClust、parentClustに代入する（ステップ２１０１，２１０６）。アイコンのクラスタと、部分木のクラスタとの

ポインタを繋ぎ替え、アイコンのクラスタを削除する処理を行なう。すなわち、parentClust.leftまたはparentClust.rightに登録されているiconClustのclusterNoをiconClust.leftに変更する（ステップ２１０７）。その後、iconClustを削除して処理を終了する（ステップ２１０８）。

【００４９】

図２２は、図１７における検索処理１７１０の詳細フローである。

選択した枝以下に対応する部分木のルートノードのクラスタのclusterNoをclustNoに代入する（ステップ２２０１）。また、部分木の先頭からのインデックスを表すleafNoを１で初期化しておく（ステップ２２０２）。またi =1,…,key_numに対して、search[i].timesを0 、search[i].placeをnullで初期化しておく（ステップ２２０３）。次に、再帰的にクラスタ木に対するトリーウォークを実行し、searchで指定したキーワードをもつ遺伝子の単語検索処理（処理A）を行なう（ステップ２２０５）。引数としてclustNo、leafNoを渡す。単語検索処理については、後で詳しく説明する。処理Aを終えると、search構造体に検索結果が入力され、処理を終了する。

【００５０】

図２３は、図２２の単語検索処理（処理A）の詳細フローである。

引数で渡されたclustNo、leafNoをそれぞれclustNo、leafNoに代入する（ステップ２３００）。また、clusterNoの指すclusterをtargetClustに代入する（ステップ２３０１）。キーワード検索のカウンタを示すi を０に設定しておく（ステップ２３０２）。

【００５１】

次に、targetCluster.typeがleafかどうかを判定する（ステップ２３０３）。leafであるとき、leafに対応する遺伝子情報とキーワード辞書ファイルから読み込んだキーワードとの比較が終わるまで、以下の処理を繰り返し行なう。すなわち、iがkey_numになるまで繰り返し行なう（ステップ２３０４）。まず、targetClust.geneIDのgeneIDに対応する遺伝子情報構造体gene_infoの属性の中に、search[i].keywordの用語が入っているか判別する（ステップ２３０５）。もし入っていたら、部分木でキーワード（search[i].keyword）が発見された回数を示すs

earch[i].timesをひとつインクリメントし、部分木での発見した位置のインデックスを示すsearch[i].placeに現在位置のleafNoを登録する（ステップ２３０７）。キーワードの検索カウンタi をひとつインクリメントし、ステップ２３０４に戻る。ステップ２３０４において、iがkey_numになったとき、即ちすべてのキーワードとの比較が終わったら、部分木のインデックスであるleafNoをひとつインクリメントし、処理を終わる（ステップ２３０９）。

【００５２】

また、ステップ２３０３において、targetCluster.typeがleafではなかった場合、子供のノードをたどる処理を行なう。targetClust.leftをclustNoに代入し（ステップ２３１０）、左の子ノードに対しclustNoとleafNoとを引数として再び単語検索処理（処理A）を行なう（ステップ２３１１）。targetCluster.typeがiconのときは、targetCluster.rightには子供ノードがないので、処理を終了する（ステップ２３１２）。ステップ２３１２において、targetCluster.typeがiconでない場合、これはnode型clusterを表す。clustNoにtargetClust.rightを代入し（ステップ２３１３）、右の子ノードに対しclustNoとleafNoとを引数として再び単語検索処理（処理A）を行ない、処理を終了する（ステップ２３１４）。

【００５３】

図２４は、図１７における部分木の遺伝子データの読み込み処理１７１５の詳細フローである。

新しく部分木を読み込んでウィンドウを作成するので、新しいウィンドウIDを示すwidをひとつインクリメントしておく（ステップ２４０１）。また、樹状図の葉の総数を表すleaf_numを0に初期化しておく（ステップ２４０２）。選択した枝以下に対応する部分木のルートノードのクラスタにおけるclusterNoをclustNoに代入する（ステップ２４０３）。最後に、部分木のleaf型クラスタに対して、新規clusterを生成する処理（処理B）を行なう（ステップ２４０４）。現在のクラスタを示すclustNoをこの処理の引数として渡す。この処理の詳細は後で説明する。すべてのleafを読み込み、leafに対応するclusterをすべて生成し処理を終了する。

【００５４】

図２５は、図２４における部分木のleafに対して新規にクラスタを生成する処理２４０４の詳細フローである。

引数で渡されたclustNoをclustNoとし、clustNoの指すclusterをtargetClustとする（ステップ２５０１，２５０２）。次に、targetCluster.typeがleafかどうかを判定する（ステップ２５０３）。leafであるならば、部分木のleafの数のカウンタであるleaf_numをひとつインクリメントする（ステップ２５０４）。次に新しいウィンドウの初期値となるleaf型クラスタ構造体を生成する。すなわち、clusterを生成し、typeにleafを、clusterNoにleaf_numを、geneIDにtargetCluster.geneIDを、windowIDにwidを登録し処理を終了する（ステップ２５０５）。

【００５５】

またステップ２５０３において、targetCluster.typeがleafではなかった場合、子供のノードをたどる処理を行なう。すなわち、targetClust.leftをclustNoに代入し（ステップ２５０６）、左の子ノードに対し、clustNoを引数として再び新規にクラスタを生成する処理（処理B）を行なう（ステップ２５０７）。targetCluster.typeがiconのときは、targetCluster.rightに子供ノードはないので、これで処理を終了する（ステップ２５０８）。ステップ２５０８において、targetCluster.typeがiconでない場合、これはnode型clusterを表している。従って、clustNoにtargetClust.rightを代入し（ステップ２５０９）、右の子ノードに対しclustNoを引数として再び新規にクラスタを生成する処理（処理B）を行い、処理を終了する（ステップ２５１０）。

以上では解析結果を表示装置画面に表示する例を説明したが、多色プリンタで印刷出力する構成であってもよい。すなわち、本発明でいう表示とは、プリンタによって視覚的に印刷出力する概念を含むものである。

【００５６】

【発明の効果】

以上示したように、本発明によると、樹状図に対して様々なクラスタリング手法を適用し、部分木をアイコン化したり、別ウィンドウで表示するなど、遺伝子

の発現解析等を支援する方法を提供することができる。

【図面の簡単な説明】

【図１】

標準的クラスタ分析結果の表示例を示す図。

【図２】

クラスタリング方法の違いの例の説明図。

【図３】

クラスタリング方法によらない樹状図の表示例を示す図。

【図４】

発現パターンが類似している遺伝子群を含む樹状図の例を示す図。

【図５】

本発明の樹状図表示システムによる画面表示例を示す図。

【図６】

本発明の樹状図表示システムによる他の画面表示例を示す図。

【図７】

本発明の樹状図表示システムによる他の画面表示例を示す図。

【図８】

本発明の樹状図表示システムによる他の画面表示例を示す図。

【図９】

本発明による樹状図表示システムの構成例を示す図。

【図１０】

遺伝子発現パターンデータの例を示す図。

【図１１】

遺伝子情報構造体の例を示す図。

【図１２】

クラスタ構造体の例を示す図。

【図１３】

クラスタ木構造の生成例を示す図。

【図１４】

クラスタ間距離を格納する配列の例を示す図。

【図１５】

各ウィンドウのルートノードを格納する配列の例を示す図。

【図１６】

検索の問合せ及び結果を格納する構造体の例を示す図。

【図１７】

本システムの概略処理フロー例を示す図。

【図１８】

遺伝子データの読み込み処理のフローを示す図。

【図１９】

クラスタ分析処理のフローを示す図。

【図２０】

クラスタ分析処理のフローを示す図。

【図２１】

（非）アイコン化処理のフローを示す図。

【図２２】

遺伝子情報を検索対象とした検索処理のフローを示す図。

【図２３】

単語検索処理（処理A）のフローをを示す図。

【図２４】

部分木の遺伝子データの読み込み処理の説明図。

【図２５】

部分木のleafに対して新規にclusterを生成する処理（処理Ｂ）の説明図。

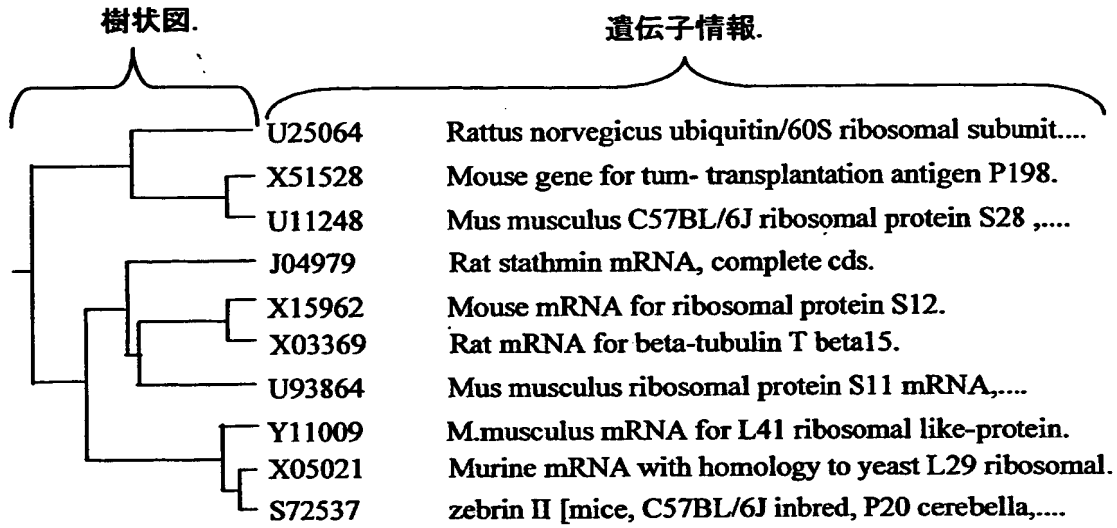
【符号の説明】

４０１…樹状図の中で発現過程が類似した遺伝子群の例、４０２…樹状図の中で発現過程が４０１の遺伝子群と大きく異なる遺伝子の例（その１）、４０３…樹状図の中で発現過程が４０１と大きく異なる遺伝子の例（その２）、５０１…クラスタリングにおける分類アルゴリズムの選択メニュー、５０２…クラスタリングにおける（非）類似度の選択メニュー、５０３…メニューウィンドウ、５０

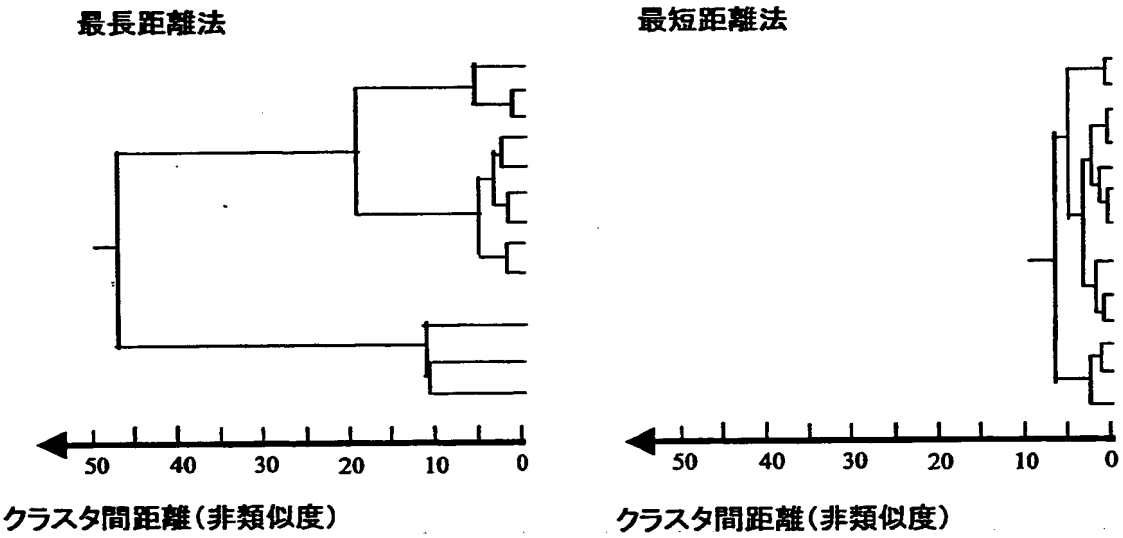
４…マウスカーソル、５０５…選択された枝（部分木）、７０１…アイコン化した部分木の例、８０１…キーワード検索結果のウィンドウ例、８０２…選択されたキーワード、８０３…遺伝子情報の中に予め定めたキーワードが含まれる遺伝子に対するマーク、８０４…マウスカーソル

【書類名】　図面

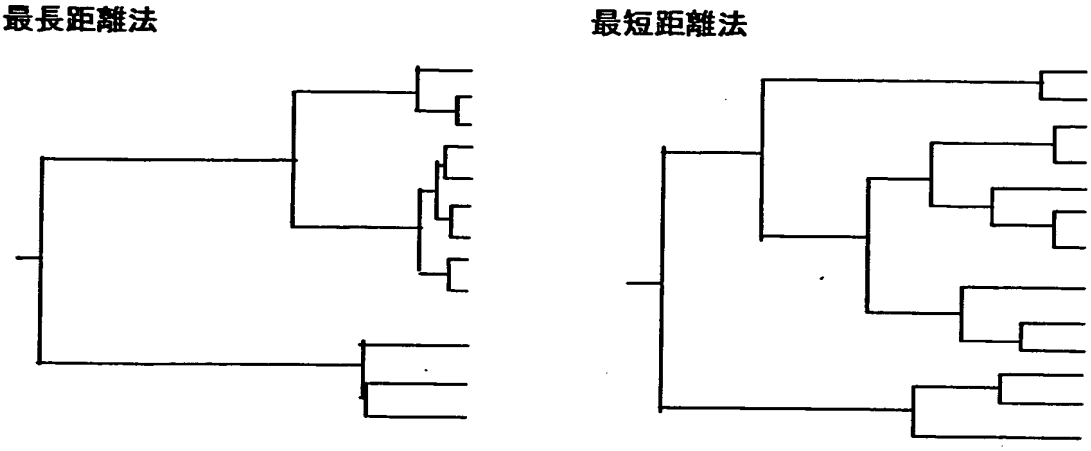
【図１】


樹状図.
遺伝子情報.
U25064
Rattus norvegicus ubiquitin/60S ribosomal subunit....
X51528
Mouse gene for tum- transplantation antigen P198.
U11248
Mus musculus C57BL/6J ribosomal protein S28 ,....
J04979
Rat stathmin mRNA, complete cds.
X15962
Mouse mRNA for ribosomal protein S12.
X03369
Rat mRNA for beta-tubulin T beta15.
U93864
Mus musculus ribosomal protein S11 mRNA,....
Y11009
M.musculus mRNA for L41 ribosomal like-protein.
X05021
Murine mRNA with homology to yeast L29 ribosomal.
S72537
zebrin II [mice, C57BL/6J inbred, P20 cerebella,....

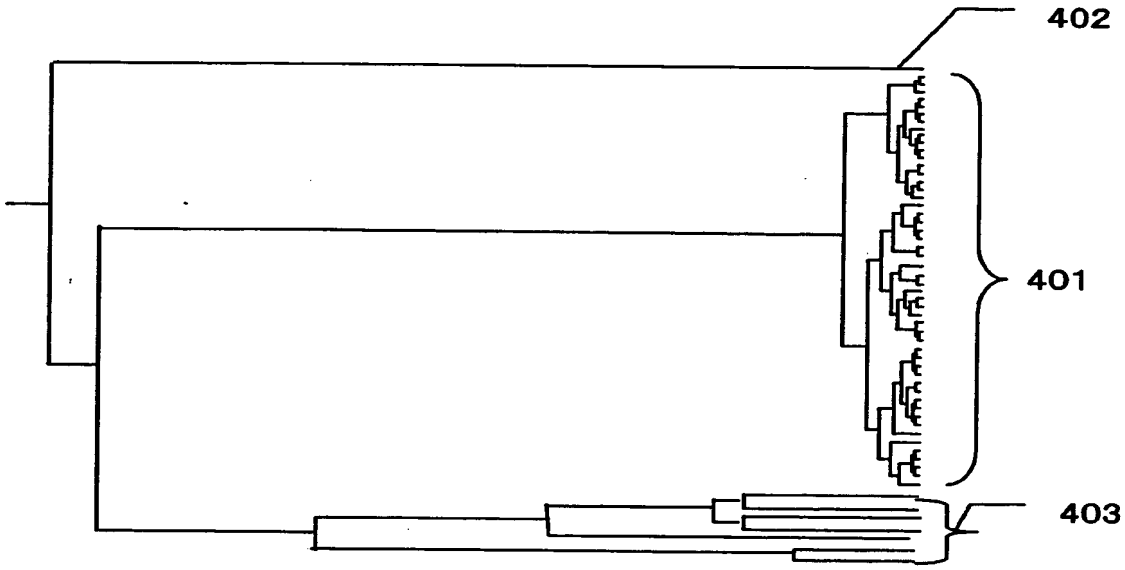
【図２】


最長距離法
50
40
30
20
10
0
クラスタ間距離(非類似度)
最短距離法
50
40
30
20
10
0
クラスタ間距離(非類似度)

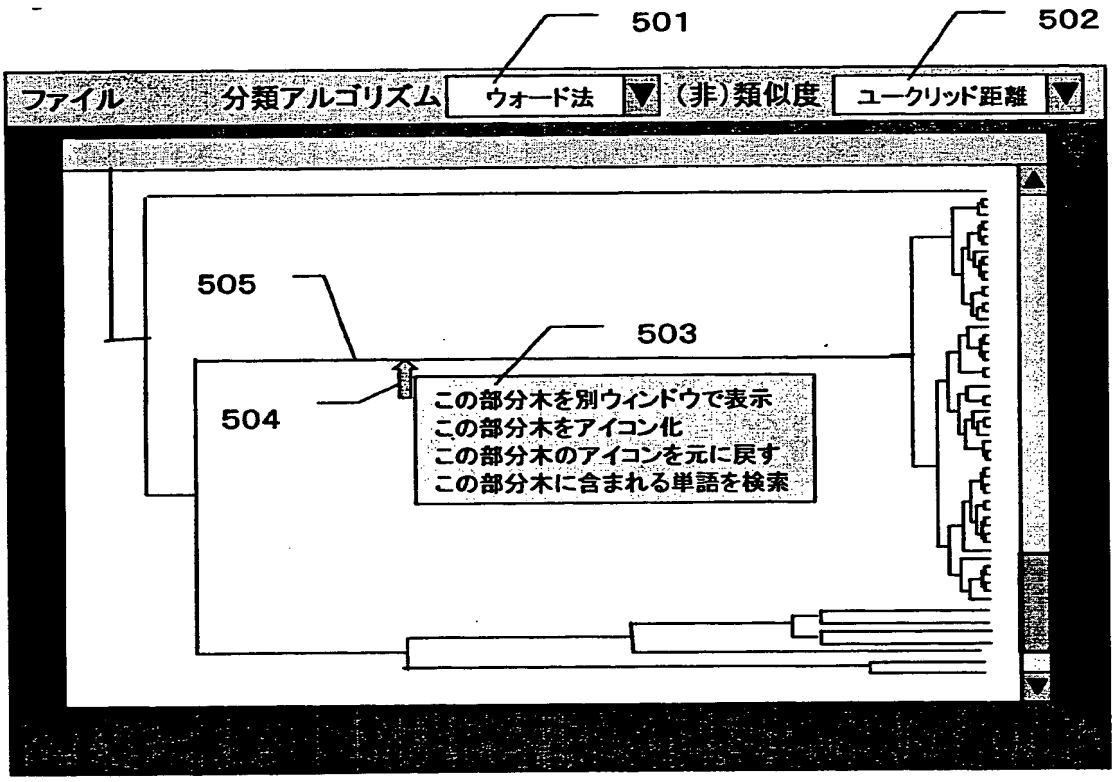
【図３】



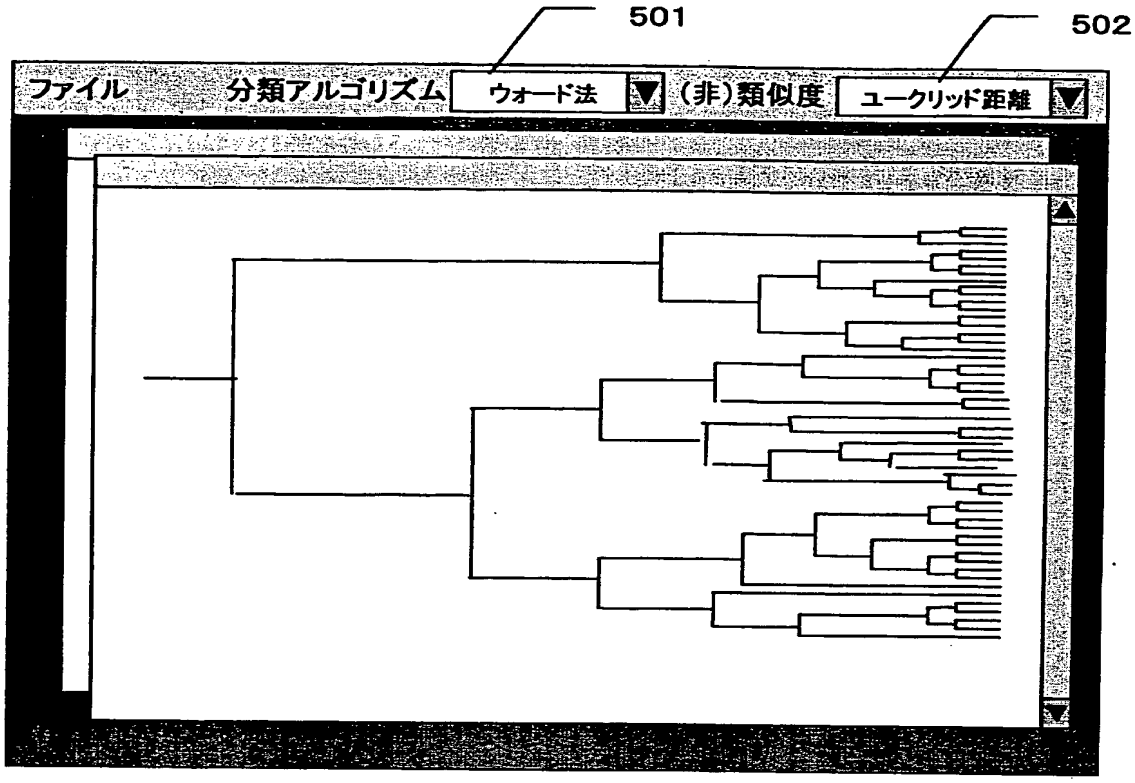
【図４】



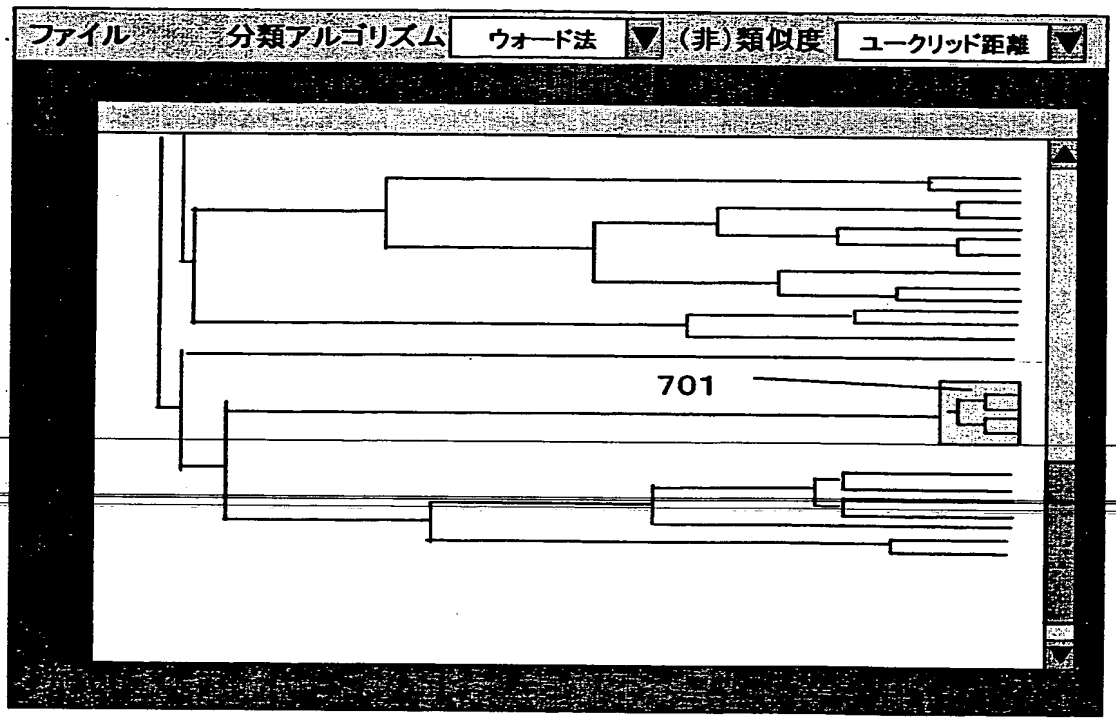
【図５】


501
502
ファイル
分類アルゴリズム
ウォード法
(非)類似度
ユークリッド距離
505
503
504
この部分木を別ウィンドウで表示
この部分木をアイコン化
この部分木のアイコンを元に戻す
この部分木に含まれる単語を検索

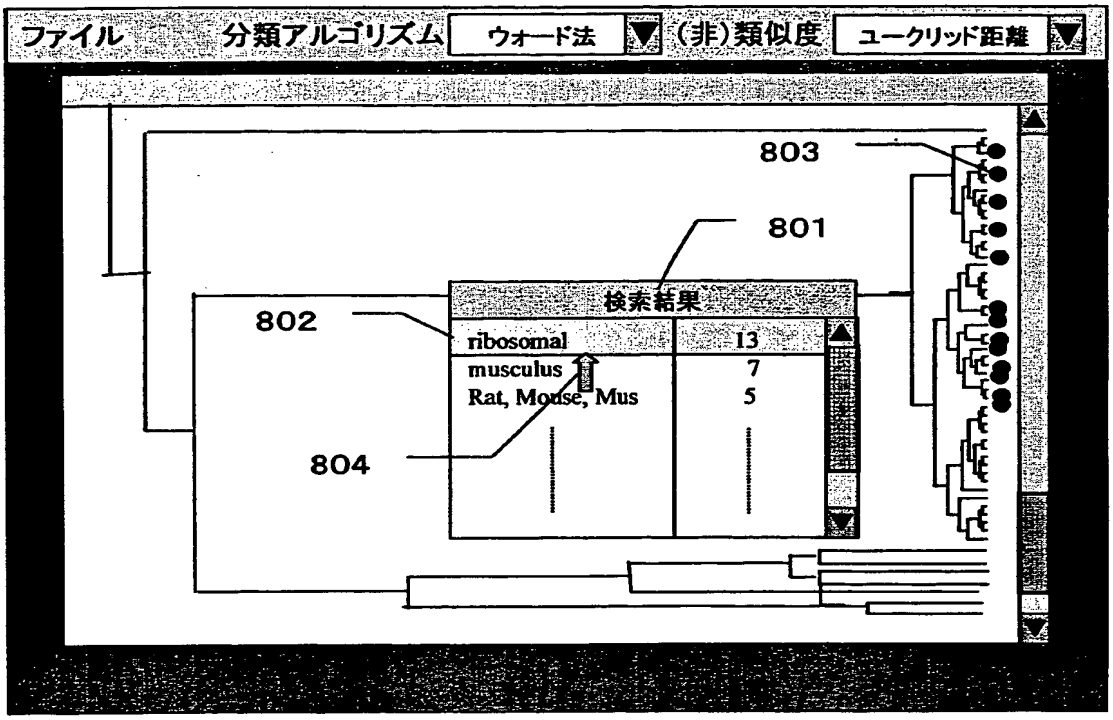
【図６】


501
502
ファイル
分類アルゴリズム
ウォード法
(非)類似度
ユークリッド距離

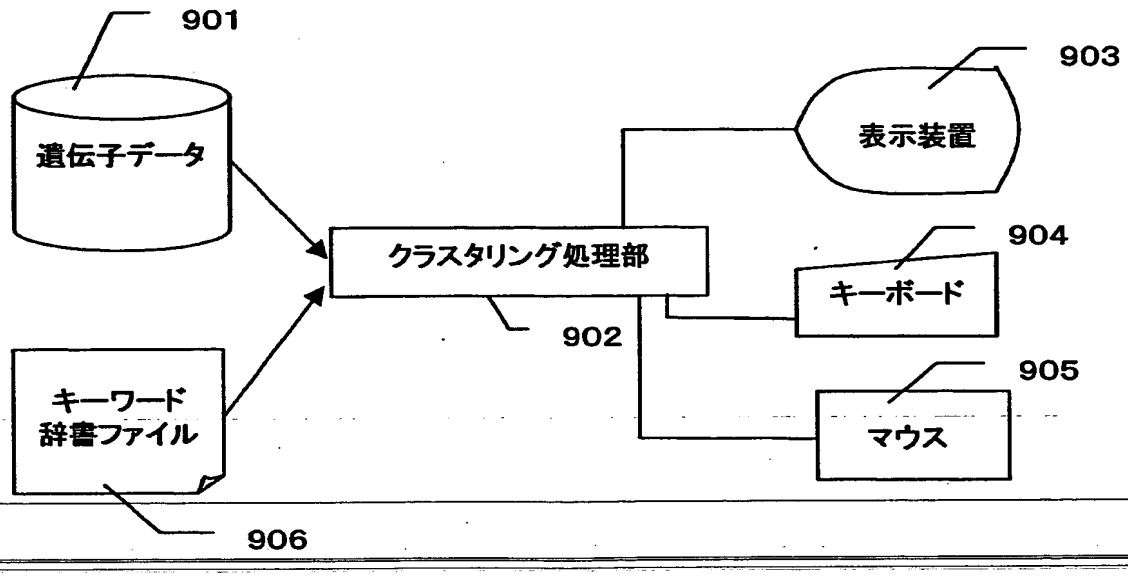
【図７】


ファイル
分類アルゴリズム
ウォード法
(非)類似度
ユークリッド距離
701

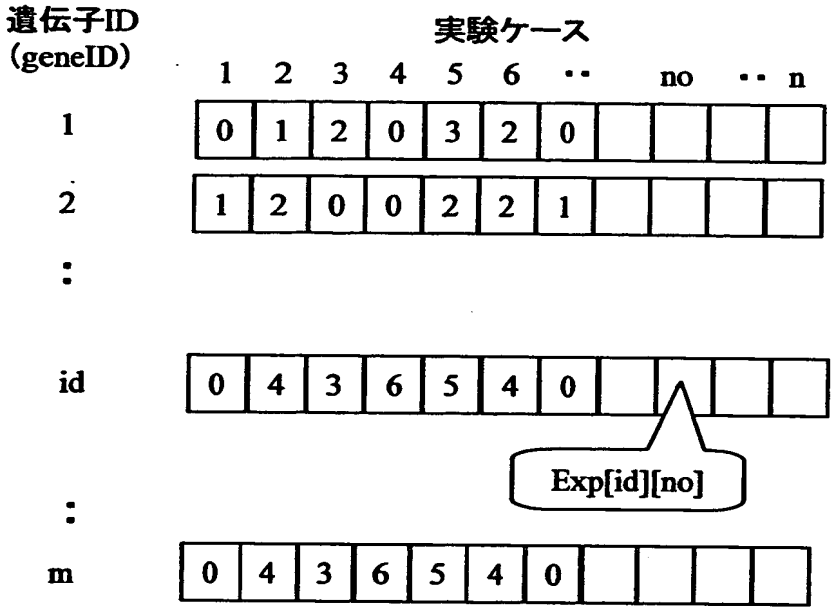
【図８】


ファイル
分類アルゴリズム
ウォード法
(非)類似度
ユークリッド距離
803
801
検索結果
802
ribosomal
13
musculus
7
Rat, Mouse, Mus
5
804

【図９】


901
遺伝子データ
903
表示装置
クラスタリング処理部
904
キーボード
902
キーワード
辞書ファイル
905
マウス
906

【図１０】



【図１１】

gene_info

	メンバ名	値
1101		
1102	geneID	17
1103	ORF	YBL084C
1104	name	ANAPHASE-PROMOTING COMPLEX SUBUNIT
	function	CELL CYCLE

【図１２】

cluster

	メンバ名	値
1201	type	leaf
1202	left	
1203	right	
1204	distance	
1205	clusterNo	17
1206	geneID	11
1207	windowID	3

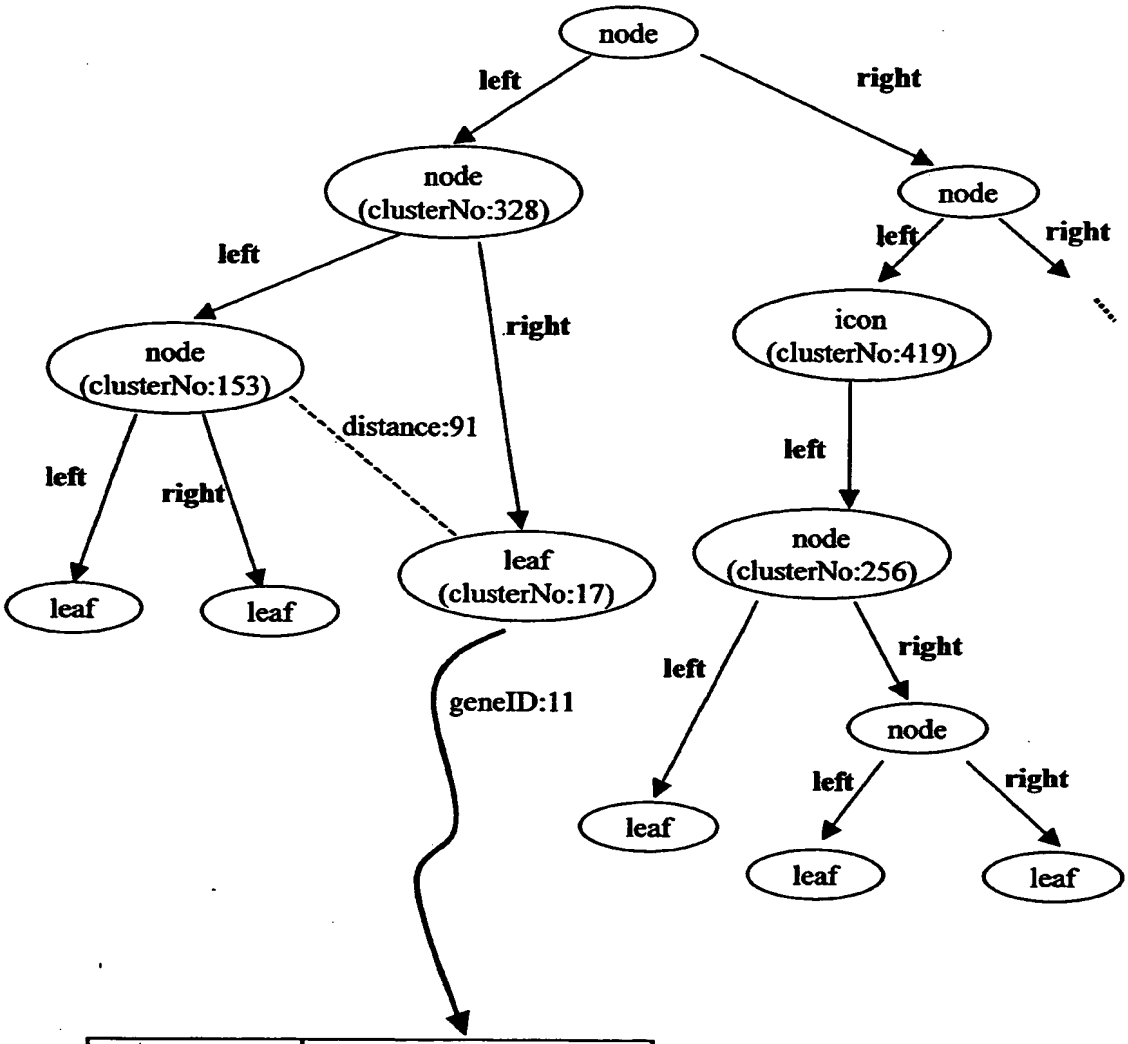
cluster

メンバ名	値
type	node
left	153
right	17
distance	91
clusterNo	328
geneID	
windowID	3

cluster

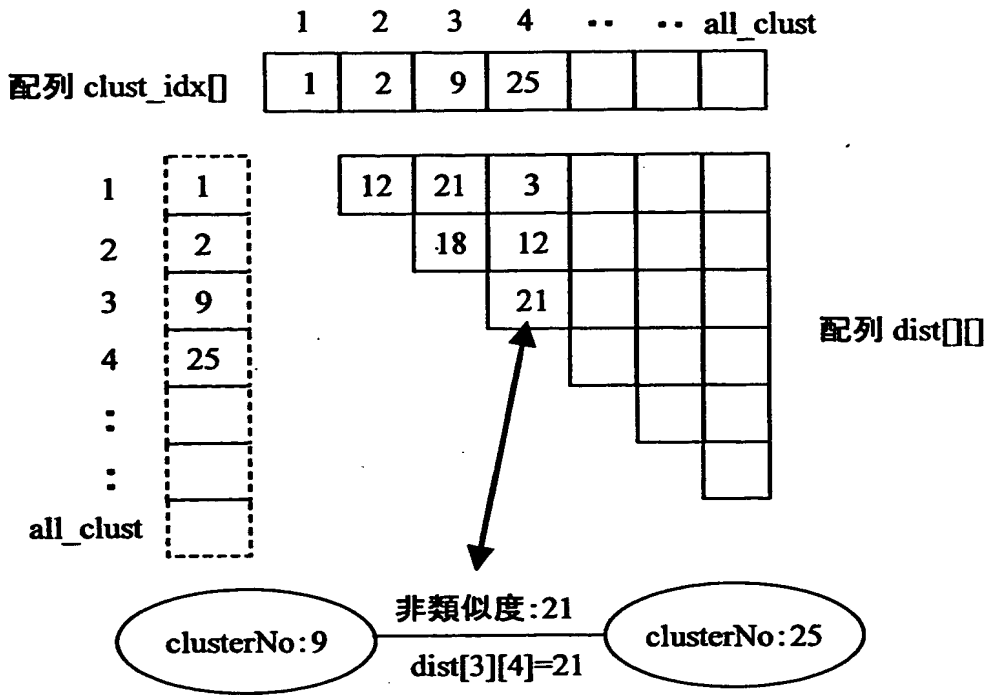
メンバ名	値
type	icon
left	256
right	
distance	
clusterNo	419
geneID	
windowID	3

【図１３】

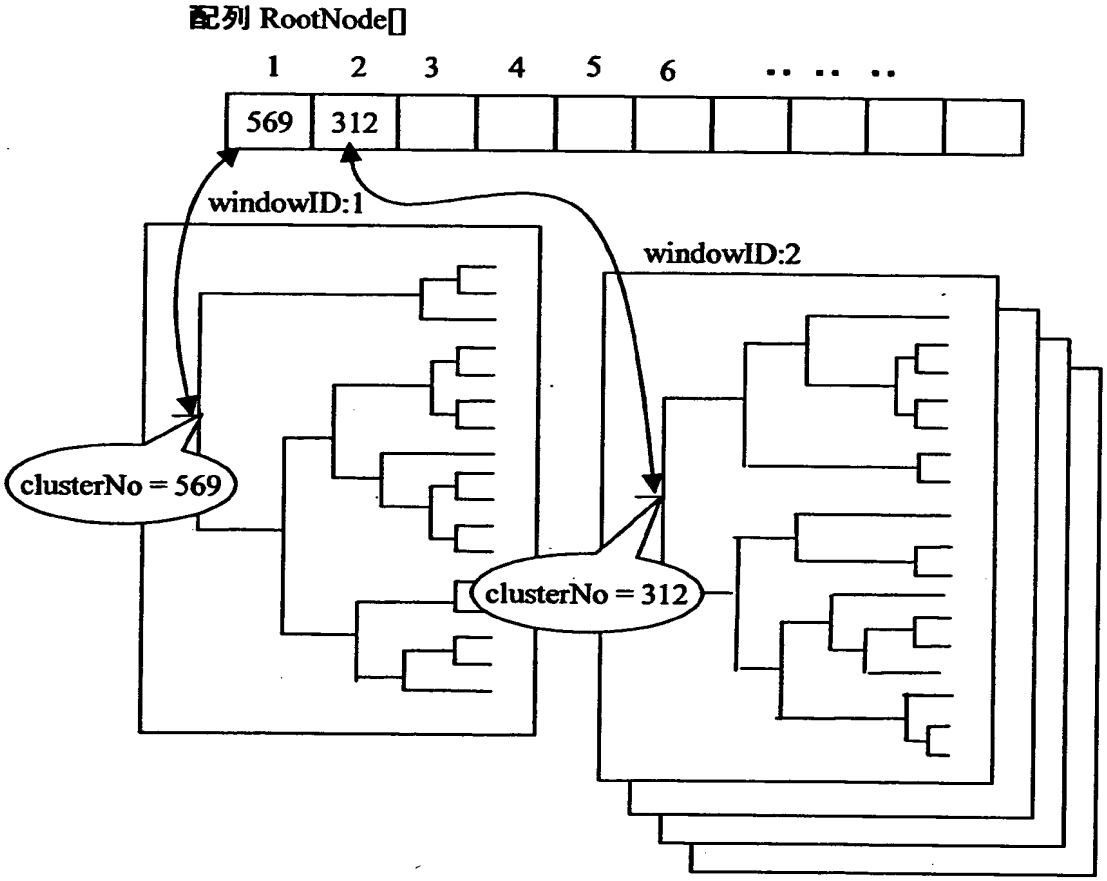


geneID	11
ORF	YDR342C
name	HEXOSE PERMEASE
function	TRANSPORT

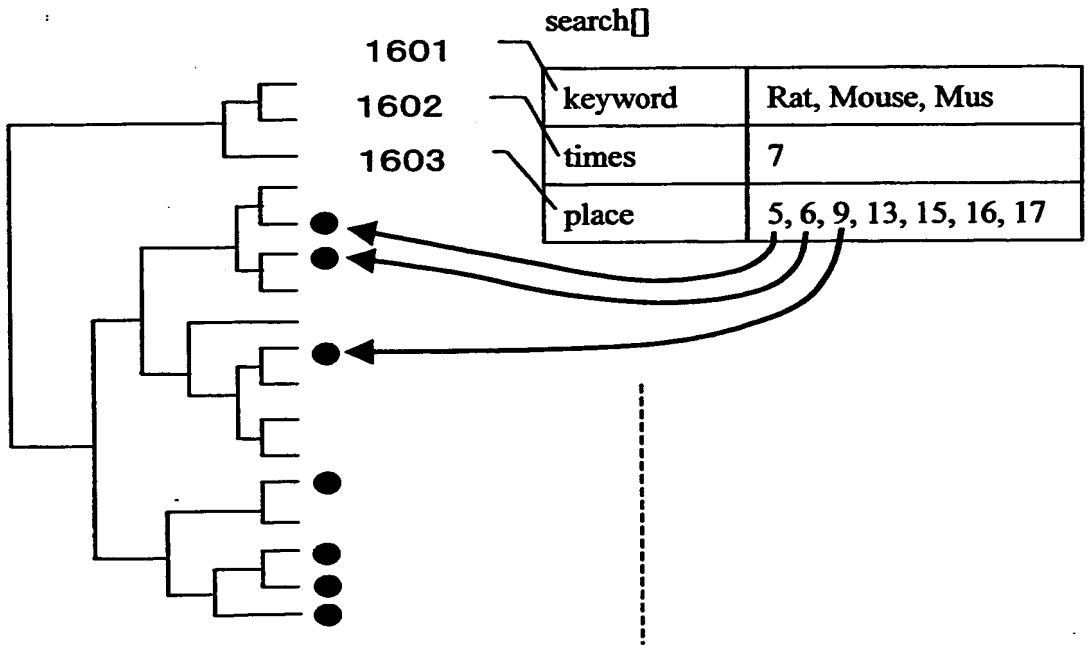
【図１４】


1 2 3 4 .. .. all_clust
配列 clust_idx[]
1 2 9 25
1
2
3
4
all_clust
1
2
9
25
12 21 3
18 12
21
配列 dist[][]
clusterNo:9
非類似度:21
dist[3][4]=21
clusterNo:25

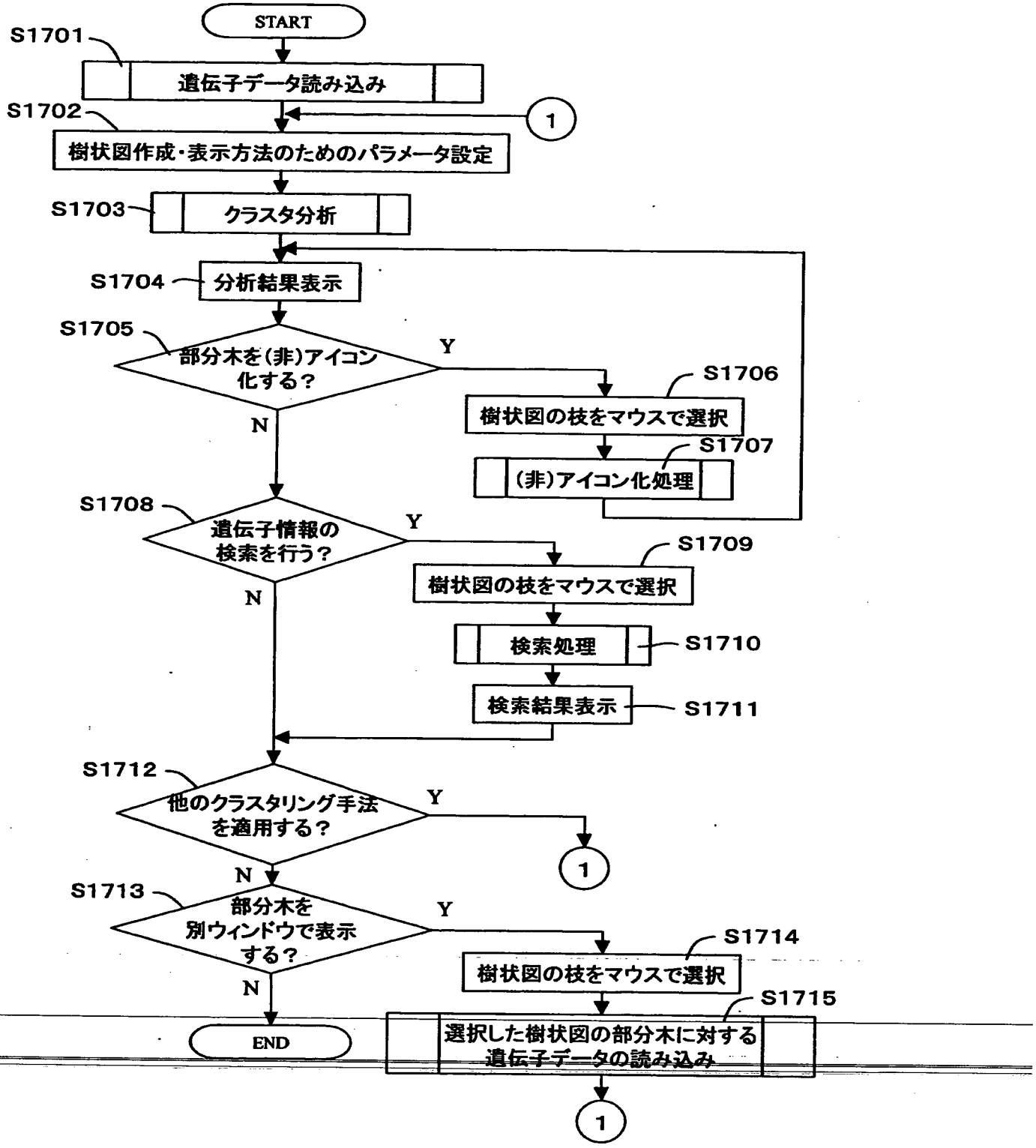
【図１５】


配列 RootNode[]
1
2
3
4
5
6
569
312
windowID:1
windowID:2
clusterNo = 569
clusterNo = 312

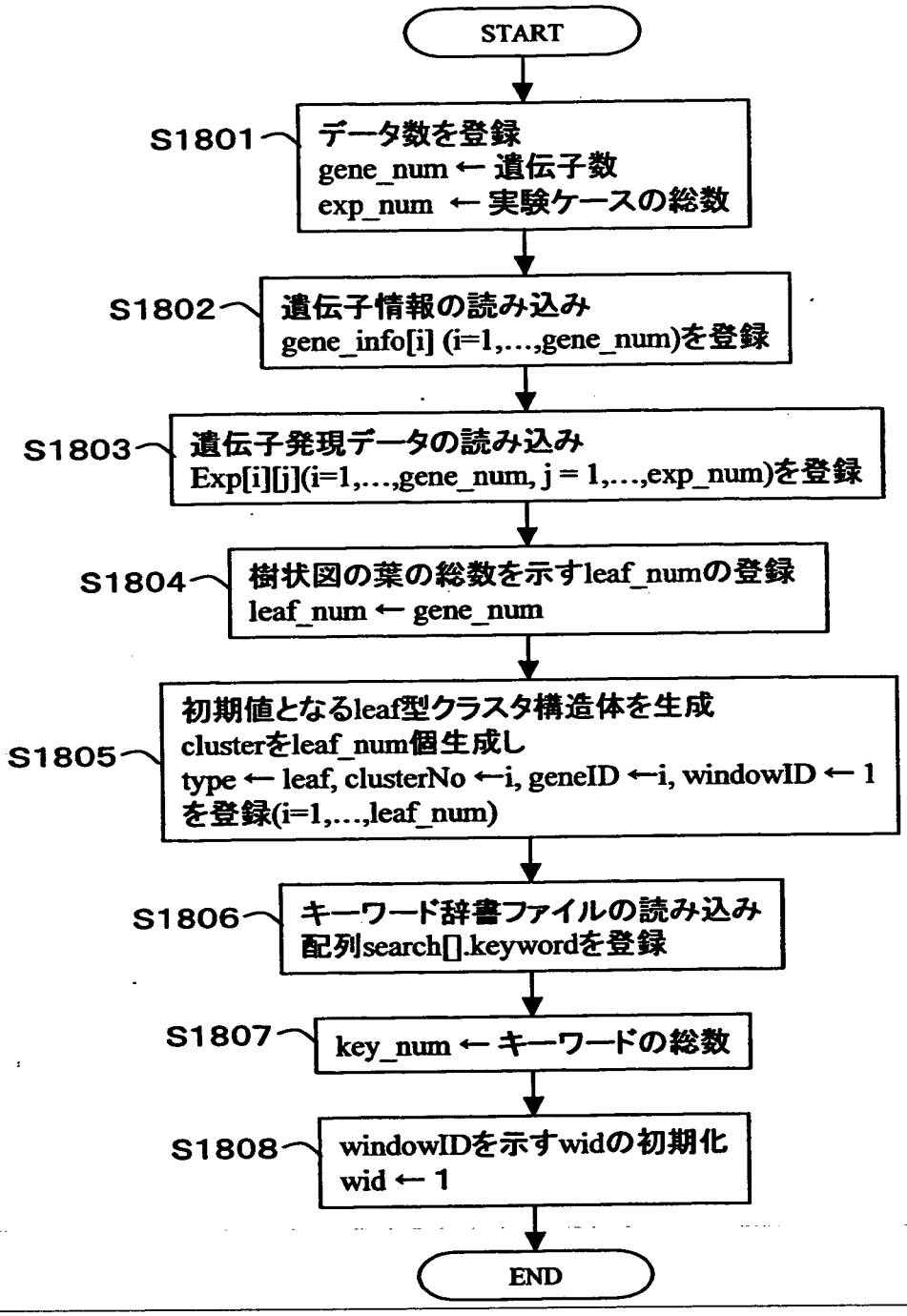
【図１６】


search[]
1601
1602
1603
keyword
Rat, Mouse, Mus
times
7
place
5, 6, 9, 13, 15, 16, 17

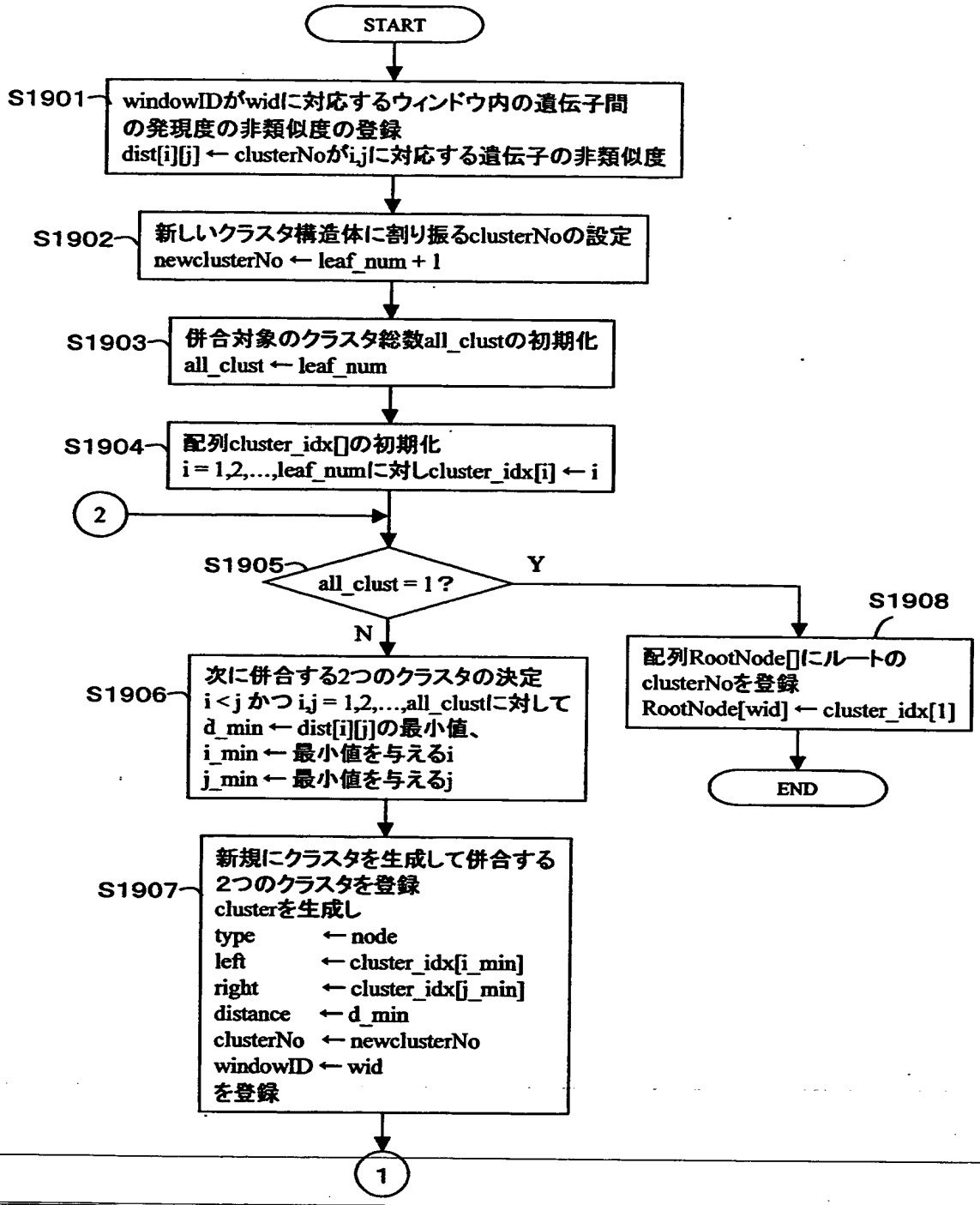
【図１７】


START
S1701
遺伝子データ読み込み
S1702
1
樹状図作成・表示方法のためのパラメータ設定
S1703
クラスタ分析
S1704
分析結果表示
S1705
部分木を(非)アイコン化する？
Y
S1706
樹状図の枝をマウスで選択
S1707
(非)アイコン化処理
N
S1708
遺伝子情報の検索を行う？
Y
S1709
樹状図の枝をマウスで選択
検索処理
S1710
検索結果表示
S1711
N
S1712
他のクラスタリング手法を適用する？
Y
1
N
S1713
部分木を別ウィンドウで表示する？
Y
S1714
樹状図の枝をマウスで選択
S1715
選択した樹状図の部分木に対する遺伝子データの読み込み
N
END
1

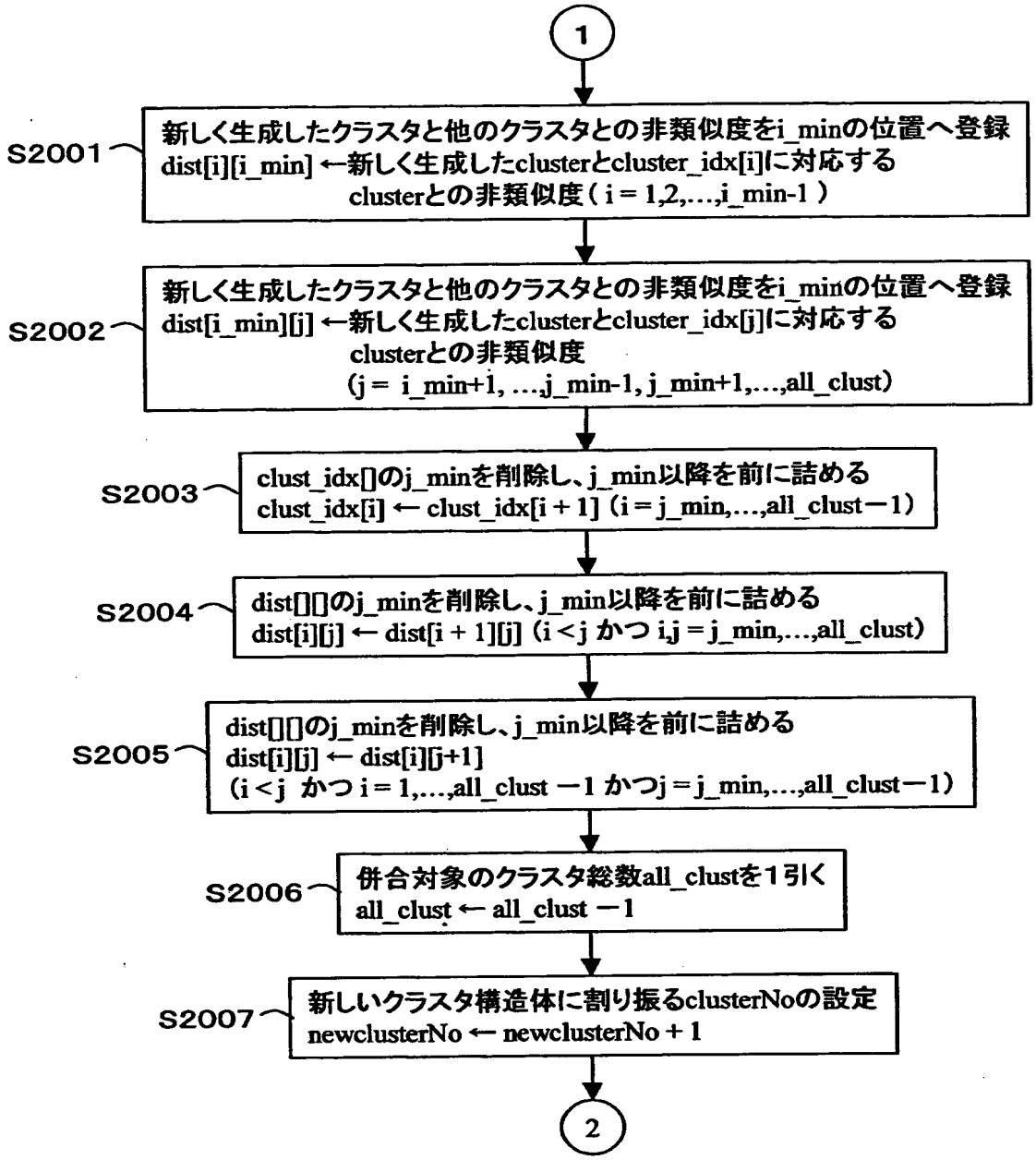
【図１８】


START
S1801
データ数を登録
gene_num ← 遺伝子数
exp_num ← 実験ケースの総数
S1802
遺伝子情報の読み込み
gene_info[i] (i=1,…,gene_num)を登録
S1803
遺伝子発現データの読み込み
Exp[i][j](i=1,…,gene_num, j = 1,…,exp_num)を登録
S1804
樹状図の葉の総数を示すleaf_numの登録
leaf_num ← gene_num
S1805
初期値となるleaf型クラスタ構造体を生成
clusterをleaf_num個生成し
type ← leaf, clusterNo ←i, geneID ←i, windowID ← 1
を登録(i=1,…,leaf_num)
S1806
キーワード辞書ファイルの読み込み
配列search[].keywordを登録
S1807
key_num ← キーワードの総数
S1808
windowIDを示すwidの初期化
wid ← 1
END

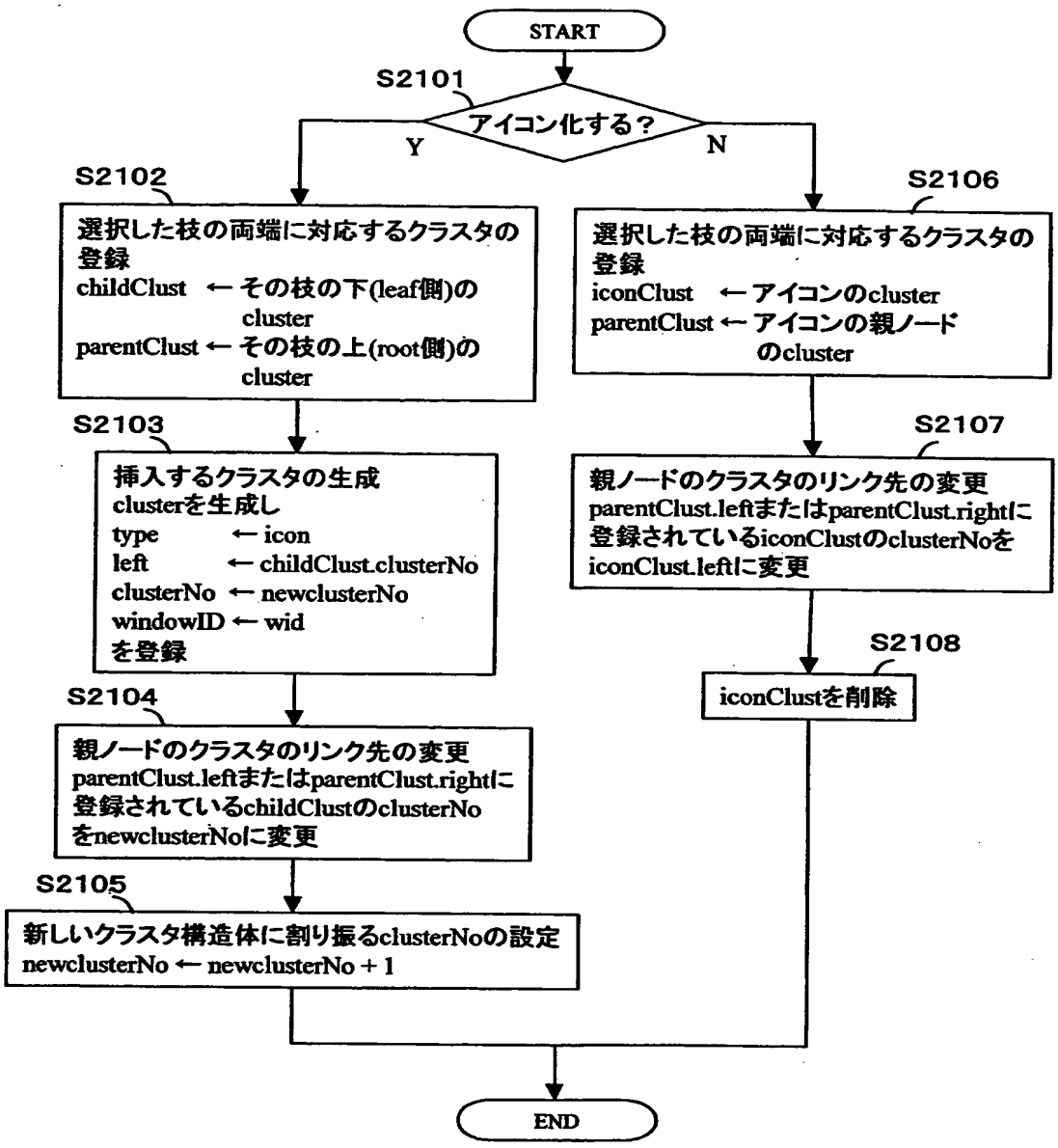
【図１９】


START
S1901
windowIDがwidに対応するウィンドウ内の遺伝子間
の発現度の非類似度の登録
dist[i][j] ← clusterNoがi,jに対応する遺伝子の非類似度
S1902
新しいクラスタ構造体に割り振るclusterNoの設定
newclusterNo ← leaf_num + 1
S1903
併合対象のクラスタ総数all_clustの初期化
all_clust ← leaf_num
S1904
配列cluster_idx[]の初期化
i = 1,2,...,leaf_numに対しcluster_idx[i] ← i
2
S1905
all_clust = 1？
Y
N
S1908
配列RootNode[]にルートの
clusterNoを登録
RootNode[wid] ← cluster_idx[1]
END
S1906
次に併合する2つのクラスタの決定
i < j かつ i,j = 1,2,...,all_clustに対して
d_min ← dist[i][j]の最小値、
i_min ← 最小値を与えるi
j_min ← 最小値を与えるj
S1907
新規にクラスタを生成して併合する
2つのクラスタを登録
clusterを生成し
type ← node
left ← cluster_idx[i_min]
right ← cluster_idx[j_min]
distance ← d_min
clusterNo ← newclusterNo
windowID ← wid
を登録
1

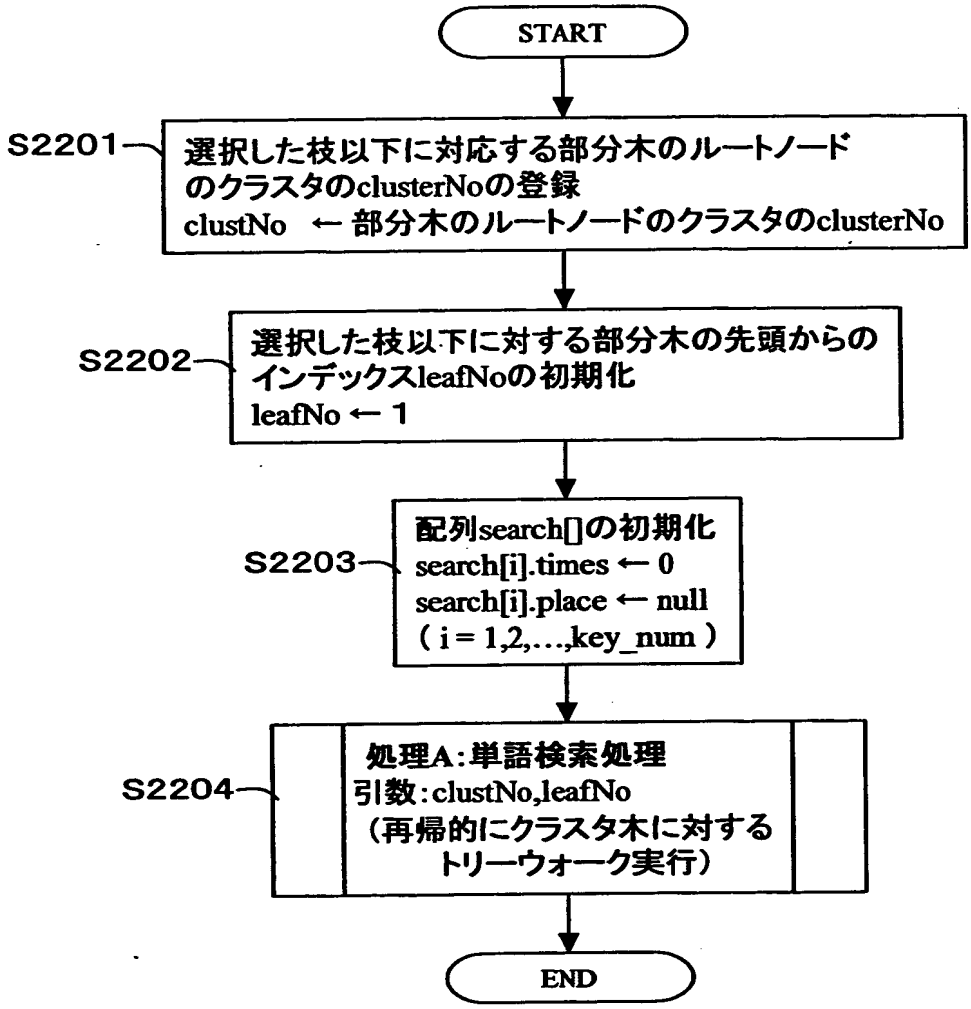
【図２０】


1
S2001
新しく生成したクラスタと他のクラスタとの非類似度をi_minの位置へ登録
dist[i][i_min] ←新しく生成したclusterとcluster_idx[i]に対応する
clusterとの非類似度（i＝1,2,…,i_min-1）
S2002
新しく生成したクラスタと他のクラスタとの非類似度をi_minの位置へ登録
dist[i_min][j] ←新しく生成したclusterとcluster_idx[j]に対応する
clusterとの非類似度
(j＝ i_min+1, …,j_min-1, j_min+1,…,all_clust)
S2003
clust_idx[]のj_minを削除し、j_min以降を前に詰める
clust_idx[i] ← clust_idx[i＋1] (i＝j_min,…,all_clust－1)
S2004
dist[][]のj_minを削除し、j_min以降を前に詰める
dist[i][j] ← dist[i＋1][j] (i＜j かつ i,j＝j_min,…,all_clust)
S2005
dist[][]のj_minを削除し、j_min以降を前に詰める
dist[i][j] ← dist[i][j+1]
(i＜j かつ i＝1,…,all_clust －1 かつj＝j_min,…,all_clust－1)
S2006
併合対象のクラスタ総数all_clustを1引く
all_clust ← all_clust －1
S2007
新しいクラスタ構造体に割り振るclusterNoの設定
newclusterNo ← newclusterNo＋1
2

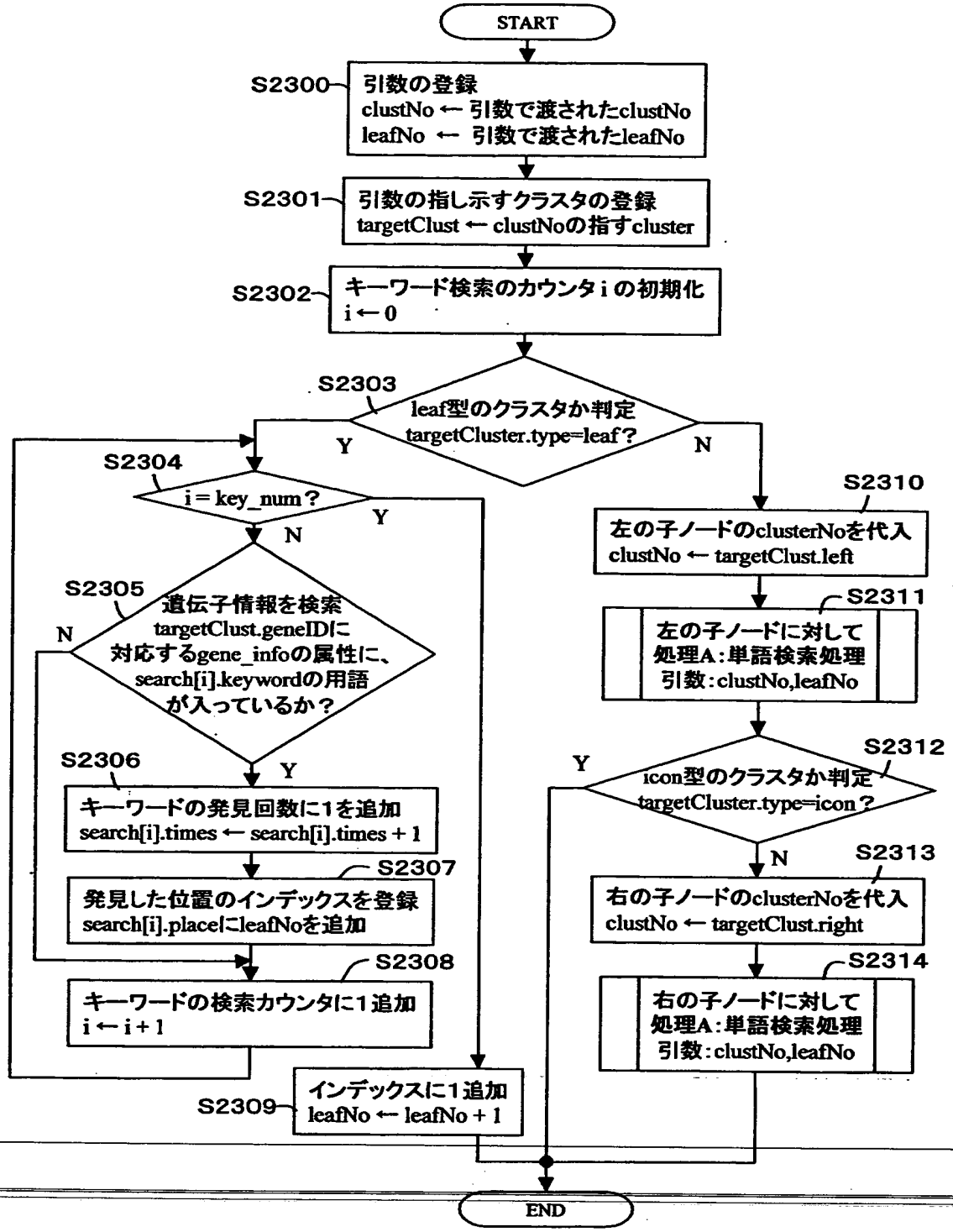
【図２１】


START
S2101
アイコン化する？
Y
N
S2102
選択した枝の両端に対応するクラスタの登録
childClust ← その枝の下(leaf側)のcluster
parentClust ← その枝の上(root側)のcluster
S2103
挿入するクラスタの生成
clusterを生成し
type ← icon
left ← childClust.clusterNo
clusterNo ← newclusterNo
windowID ← wid
を登録
S2104
親ノードのクラスタのリンク先の変更
parentClust.leftまたはparentClust.rightに登録されているchildClustのclusterNoをnewclusterNoに変更
S2105
新しいクラスタ構造体に割り振るclusterNoの設定
newclusterNo ← newclusterNo + 1
S2106
選択した枝の両端に対応するクラスタの登録
iconClust ← アイコンのcluster
parentClust ← アイコンの親ノードのcluster
S2107
親ノードのクラスタのリンク先の変更
parentClust.leftまたはparentClust.rightに登録されているiconClustのclusterNoをiconClust.leftに変更
S2108
iconClustを削除
END

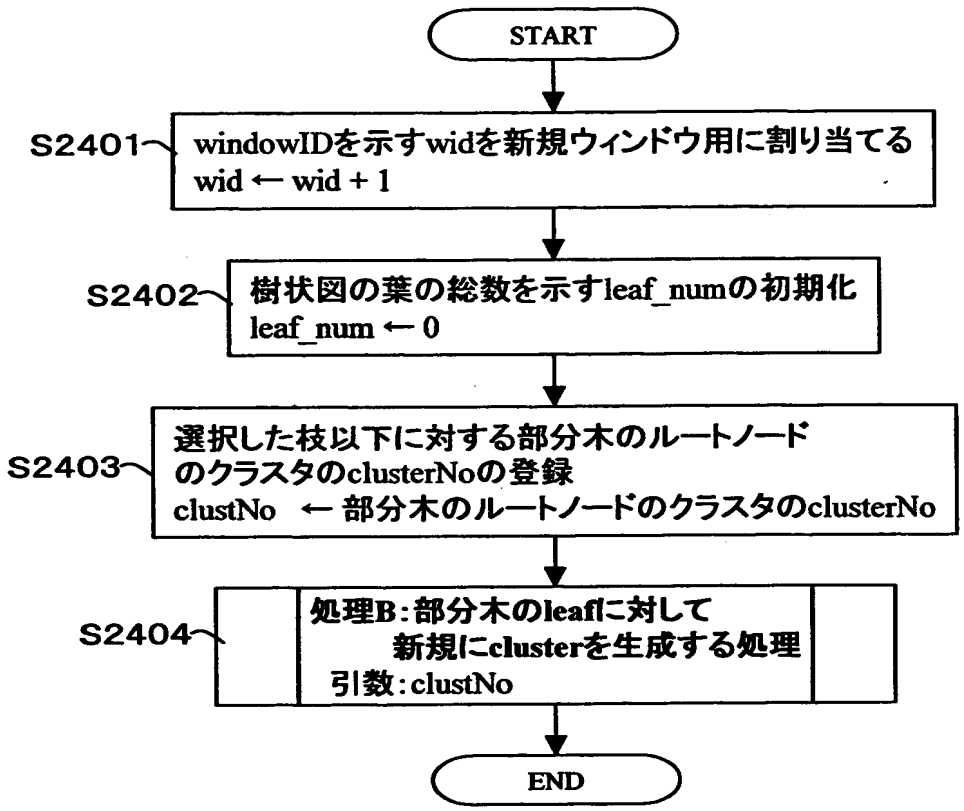
【図２２】


START
S2201
選択した枝以下に対応する部分木のルートノード
のクラスタのclusterNoの登録
clustNo ← 部分木のルートノードのクラスタのclusterNo
S2202
選択した枝以下に対する部分木の先頭からの
インデックスleafNoの初期化
leafNo ← 1
S2203
配列search[]の初期化
search[i].times ← 0
search[i].place ← null
(i = 1,2,…,key_num)
S2204
処理A:単語検索処理
引数:clustNo,leafNo
(再帰的にクラスタ木に対する
トリーウォーク実行)
END

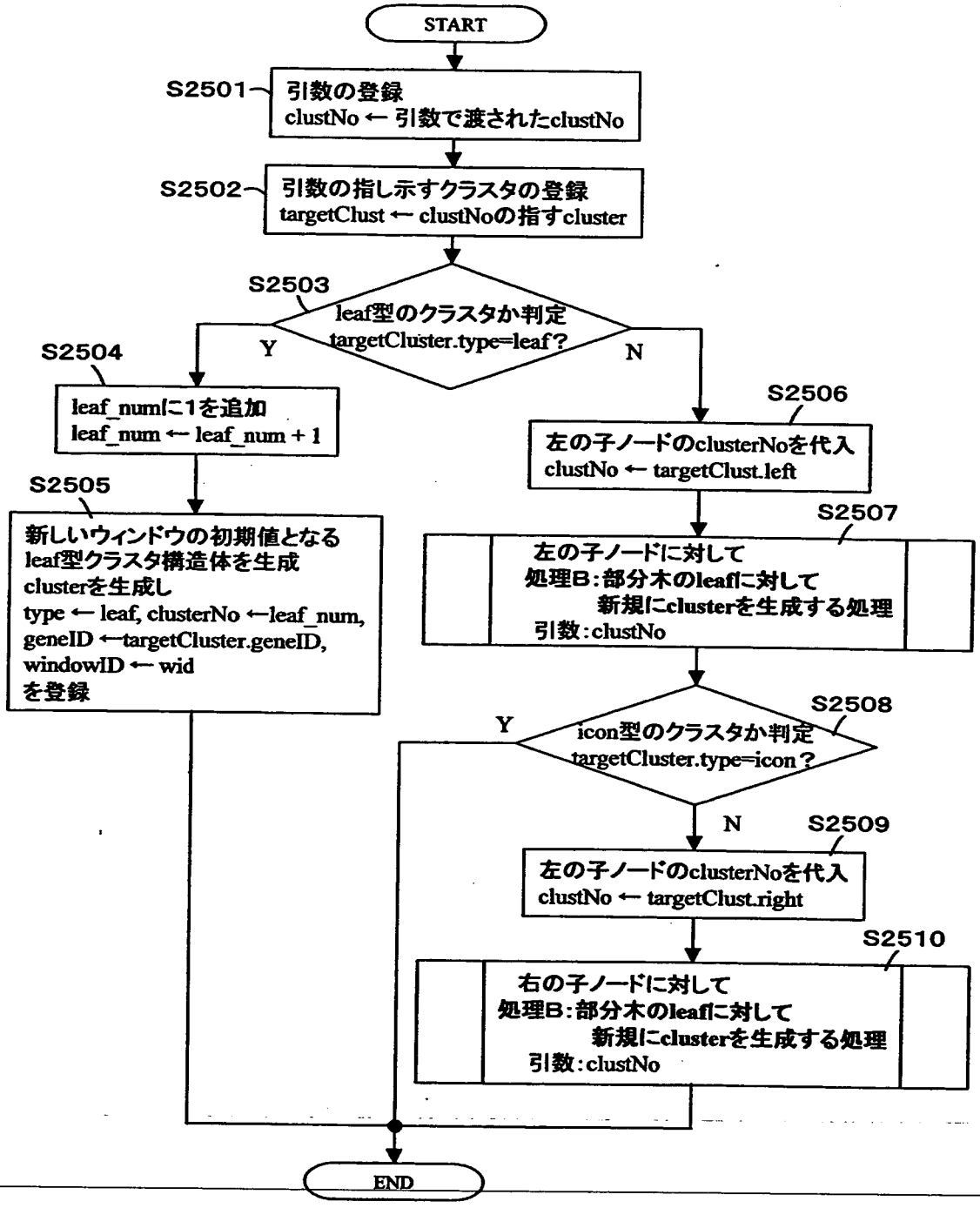
【図２３】


START
S2300
引数の登録
clustNo ← 引数で渡されたclustNo
leafNo ← 引数で渡されたleafNo
S2301
引数の指し示すクラスタの登録
targetClust ← clustNoの指すcluster
S2302
キーワード検索のカウンタ i の初期化
i ← 0
S2303
leaf型のクラスタか判定
targetCluster.type=leaf？
Y
N
S2304
i = key_num？
Y
N
S2305
遺伝子情報を検索
targetClust.geneIDに
対応するgene_infoの属性に、
search[i].keywordの用語
が入っているか？
N
Y
S2306
キーワードの発見回数に1を追加
search[i].times ← search[i].times + 1
S2307
発見した位置のインデックスを登録
search[i].placeにleafNoを追加
S2308
キーワードの検索カウンタに1追加
i ← i + 1
S2309
インデックスに1追加
leafNo ← leafNo + 1
S2310
左の子ノードのclusterNoを代入
clustNo ← targetClust.left
S2311
左の子ノードに対して
処理A：単語検索処理
引数：clustNo,leafNo
S2312
icon型のクラスタか判定
targetCluster.type=icon？
Y
N
S2313
右の子ノードのclusterNoを代入
clustNo ← targetClust.right
S2314
右の子ノードに対して
処理A：単語検索処理
引数：clustNo,leafNo
END

【図２４】


START
S2401
windowIDを示すwidを新規ウィンドウ用に割り当てる
wid ← wid + 1
S2402
樹状図の葉の総数を示すleaf_numの初期化
leaf_num ← 0
S2403
選択した枝以下に対する部分木のルートノード
のクラスタのclusterNoの登録
clustNo ← 部分木のルートノードのクラスタのclusterNo
S2404
処理B:部分木のleafに対して
新規にclusterを生成する処理
引数:clustNo
END

【図２５】


START
S2501
引数の登録
clustNo ← 引数で渡されたclustNo
S2502
引数の指し示すクラスタの登録
targetClust ← clustNoの指すcluster
S2503
leaf型のクラスタか判定
targetCluster.type=leaf？
Y
N
S2504
leaf_numに1を追加
leaf_num ← leaf_num + 1
S2505
新しいウィンドウの初期値となる
leaf型クラスタ構造体を生成
clusterを生成し
type ← leaf, clusterNo ←leaf_num,
geneID ←targetCluster.geneID,
windowID ← wid
を登録
S2506
左の子ノードのclusterNoを代入
clustNo ← targetClust.left
S2507
左の子ノードに対して
処理B:部分木のleafに対して
新規にclusterを生成する処理
引数:clustNo
S2508
Y
icon型のクラスタか判定
targetCluster.type=icon？
N
S2509
左の子ノードのclusterNoを代入
clustNo ← targetClust.right
S2510
右の子ノードに対して
処理B:部分木のleafに対して
新規にclusterを生成する処理
引数:clustNo
END

【書類名】　要約書

【要約】

【課題】　樹状図全体の枝の状態を大域的に把握すると共に個々の部分木の状態を詳細に知り、分類の絞り込みや、クラスタリング方法の選択の支援をする。

【解決手段】　樹状図の枝を選択し、選択した枝から葉の部分木に対して、別の表示ウィンドウで表示する機能、アイコン化する機能、アイコン化したものを元に戻す機能、部分木に含まれるキーワードを収集し表示する機能を備える。

【選択図】　図５

出 願 人 履 歴 情 報

識別番号　〔０００２３３０５５〕

１．変更年月日　１９９０年　８月　７日
［変更理由］　新規登録
住　所　神奈川県横浜市中区尾上町６丁目８１番地
氏　名　日立ソフトウエアエンジニアリング株式会社